



ABO: Interactive Journal for Women in the Arts, 1640-1830

Volume 5
Issue 2 *Volume 5.2 (Fall 2015)*

Article 1

2015

Chawton Novels Online, Women's Writing 1751-1834 and Computer-Aided Textual Analysis

Anne Bandry-Scubbi
University of Strasbourg, bandry@unistra.fr

Follow this and additional works at: <https://scholarcommons.usf.edu/abo>



Part of the [Digital Humanities Commons](#), [Feminist, Gender, and Sexuality Studies Commons](#), and the [Literature in English, British Isles Commons](#)

Recommended Citation

Bandry-Scubbi, Anne (2015) "Chawton Novels Online, Women's Writing 1751-1834 and Computer-Aided Textual Analysis," *ABO: Interactive Journal for Women in the Arts, 1640-1830*: Vol. 5 : Iss. 2 , Article 1.
<https://www.doi.org/http://dx.doi.org/10.5038/2157-7129.5.2.1>
Available at: <https://scholarcommons.usf.edu/abo/vol5/iss2/1>

This Digital Humanities is brought to you for free and open access by Scholar Commons. It has been accepted for inclusion in ABO: Interactive Journal for Women in the Arts, 1640-1830 by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Chawton Novels Online, Women's Writing 1751-1834 and Computer-Aided Textual Analysis

Abstract

Using Chawton House Library's "[Novels Online](#)," several corpora have been set up for a computer-aided textual analysis of the use of vocabulary by women writing "domestic novels" from 1752 to 1834. This corpus stylistics approach makes it possible to map texts according to their word usage and to identify quantitative keywords which provide vocabulary profiles through comparison and contrast with contemporary male and female canonical texts. Items identified include pronouns, markers of dialogue and of intensity; others can be grouped into specific lexical fields such as feelings. One text from the collection then forms the object of a case-study to explore a paradox: although Jane Taylor's use of vocabulary in her 1817 *Rachel* appears the most representative of the corpus made up of 42 novels by women, this Chawton text has been called "a highly original tale." Methodology and findings are both presented to address the challenge of identifying features which constitute typicality.

Keywords

Chawton novels online, Women's writing 1751-1834, Computer-aided textual analysis, Corpus stylistics, quantitative stylistics, Jane Taylor

Creative Commons License



This work is licensed under a [Creative Commons Attribution-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nd/3.0/).

Chawton “Novels Online,” Women’s Writing 1751-1834 and Computer-aided Textual Analysis

1. Chawton Novels Online and Strategies of Writing

1.1 The scope of the study

Chawton House Library’s [Novels Online](#) offers some sixty non-canonical texts of fiction in electronic form, which can be converted from PDF into machine-readable text necessary for use with software, rather than as facsimiles provided by databases like *Eighteenth Century Collections Online* (ECCO). These electronic texts all derive from the rarest books the library owns and, with two exceptions, range from 1760 to 1830.¹ Most of them are novels centered on a heroine or a small community, set in the British Isles, akin to those of Jane Austen, Chawton’s star (her house in the village is a [museum](#) and Chawton House, formerly in the possession of the Knight family, is now a library and study center focusing “on works written by women in English during the period 1600 to 1830” ([Chawton House Library](#))). Because the Chawton Novels Online have been carefully typed and checked by volunteers, their quality is high, contrary to many freely accessible electronic texts (Bandry-Scubbi, “Chawton House”). From the fifty-four which were available when this study began with a visiting fellowship at Chawton House Library in April 2013, thirty-four have been set up into a corpus (the criteria for selection are given in section 2.1 below).

These texts are used here to examine writing by women in the period from the perspective of computational stylistics, as part of an ongoing personal project entitled “Strategies of Writing.” Previous computer-aided analyses of Austen and Frances Burney (Bandry-Scubbi, “Space and Emotions,” “Evelina, Lydia, Isabella,” and “Yes, Novels”) illustrated the importance of researching these authors’ novels within the context of contemporary fiction. This project therefore relies on a corpus-based approach to assess the features of what Marilyn Butler calls “the ‘feminine’ novel—domestic comedy, centring on a heroine, in which the critical action is an inward progress towards judgment” (145) and what James Raven more bluntly describes as “domestic dramas where heroines blush, swoon, or face unbearable social ostracism because of minor breaches of decorum” (28). In this study, canonical and non-canonical novels are given the same importance, in the interest of contributing to the “more flexible kind of literary history” Isobel Grundy advocates by bringing to light both common and original writing strategies, the literary and cultural codes which canonical texts both use and transcend (“Chawton House” 179).

In the present study, the comparison and contrast of these novels among themselves and with contemporary texts help to identify the use of vocabulary by women writers of domestic fiction over the 1751-1834 period. This is looked at in a global way in sections 2 to 4, while section 5 focuses specifically on Chawton novels and section 6 consists of a case-study, zooming in on one text from the collection, Jane Taylor’s *Rachel*. Section 7 zooms out to draw conclusions on some of the writing strategies at work in the genre and the advantages of using popular literature as a benchmark in a corpus-based approach.

Before addressing these texts and approaches, however, the last element of my title, “computer-aided textual analysis,” needs to be elucidated.

1.2 Computer-aided textual analysis

Scholars working on domains covered by the collections at Chawton House Library are usually aware of John Burrows' 1987 *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. It was a forerunner in quantitative textual studies, less technical than, for instance, Louis T. Milic's *A Quantitative Approach to the Style of Jonathan Swift*, Frederick Mosteller and David Wallace's *Inference and Disputed Authorship: the Federalist* or Charles Muller's *Essai de statistique lexicale: L'illusion comique de Pierre Corneille*. All of these studies focus on high frequency words and devised innovative and inspiring methods of analysis. Their explanations of experimental methodology as well as their insights into textual genres, categories, and styles continue to prevail in recent research. Under the aegis of Franco Moretti and Mark Algee-Hewitt, the [Stanford Literary Lab](#) "applies computational criticism, in all its forms, to the study of literature," which Matthew L. Jockers puts into perspective and exemplifies in *Macroanalysis: Digital Methods & Literary History*. In the recently named field of "corpus stylistics," European corpus linguists work on literary texts, with dedicated collections published by [Routledge](#), Continuum, and [Honoré Champion](#). These include Michaela Mahlberg's 2013 *Corpus Stylistics and Dickens's Fiction* and, even closer to the present study, Bettina Fischer-Starcke's 2010 *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries*. Written by linguists, in the tradition of *Style in Fiction* by Geoffrey N. Leech and Michael H. Short, both Mahlberg's and Fischer-Starcke's studies confirm hypotheses explored by literary scholars, thereby proving their own validity, and they each bring to light "interpretative innovations" (Fischer-Starcke 199). As Burrows points out in his 1997 chapter on Austen's style, "computational methods of analysis [...] give an unprecedented access to many small touches that good readers recognize but can seldom quite define" ("Style" 186).²

The development of personal computer software since the 1990s has made such methodologies accessible to literary scholars. Many possibilities are now offered online for free ([Voyant](#), [Antconc](#), [TXM](#), etc.), with more or less demanding learning processes. Some merely produce word clouds now billowing all over the webosphere but providing little more than elegant visual short-cuts. Others provide the researcher with heuristic graphs, concordances, wordlists, and statistics.³ Such quantitative results make sense when combined with a qualitative analysis: computer-aided textual analysis requires a constant back and forth movement between the data and the text. The present study largely relies on [Hyperbase](#), a program developed in France by Etienne Brunet in 1989 and now in its tenth version, and to a lesser extent on its recent (but less user-friendly because still experimental) offspring, [TXM](#). Brunet's approach, like the ones mentioned above, enables the user to uncover stylistic "traits of contrast" (Rastier 241) by providing lists of salient words and what I find to be useful and accessible "textual imaging," that is, visual representations of data from the corpus (and no coding is necessary). In Mahlberg's words, "what corpus stylistics can do beyond the obvious provision of quantitative data, is help with the analysis of an individual text by providing various options for the comparison of one text with groups of other texts to identify tendencies, intertextual relationships, or reflections of social and cultural contexts" ("Corpus Stylistics" 221).⁴

Unlike Moretti, I am not looking for "abstract models" with which to map large swathes of literary history, including "the lost 99.5% of the archive" (*Distant Reading* 77), but rather, with John Burrows, "declare myself, first and last, a student of English literature," having taken up

“computational stylistics” (“Never Say Always Again” 13) because the methods and tools it provides make it possible to view texts from a somewhat different perspective, combining distant and close reading. David A. Brewer has recently questioned the use of “the quantitative turn” and he advocates taking into account the “qualitative difference” between “super-successful texts” which left “a huge footprint” and texts which only had one or two editions, now made available by ECCO and other such repositories (161, 166).⁵ My position is to try to get the best of both worlds by using little-known texts to trace the writing strategies of a genre. What is proposed here is the basis for a larger work in progress.

2. The Corpora

Contrasting texts or parts of texts requires setting them up into a corpus, or better still, several corpora.⁶ A corpus is a construct, and this is particularly important for literary analysis, as the integrity of texts plays an important role. A corpus can be defined as “a collection of (1) machine-readable (2) authentic texts (3) sampled to be (4) representative of a particular language or language variety” (Ho 6). I’ve learned from experience that corpora of varying homogeneity are the most useful in computer-aided textual analysis: some of mine comprise the texts by a single author (Austen, Burney, Haywood, Smollett and Sterne at this point), others span a period. This is the case of the three described below, which were set up for this study.⁷ The large amount of text partly makes up for the inevitable arbitrariness of choice. In all three corpora, the texts are arranged in chronological order.

2.1 The Chawton corpus

From the fifty-four texts available in April 2013 as “[Novels Online](#),” I have selected the thirty-four which were the most congruent with “the ‘feminine’ novel” or “domestic drama” according to the definitions by Butler and Raven quoted in section 1.1.

Although the line was occasionally difficult to draw, I have discarded the texts whose action does not take place mainly in the British Isles (unless foreign locations play a very minor role in the story), so that vocabulary linked to a foreign context does not create undue distortion.⁸ After this elimination, only two texts remained whose title claimed their status as historical fiction, and these were also omitted as they were too few to make a relevant group in the corpus.⁹ Very short or fragmented texts were not kept as both size and narrative consistency play a role in the distribution of words.¹⁰

The thirty-four texts constituting the corpus “CHAWTN34” add up to around 3.3 million words.¹¹ Fourteen of these novels were published between 1769 and 1799, and twenty between 1800 and 1830, so that 38% of the Chawton corpus dates from before 1800.¹²

Seven texts in the corpus were issued by the (in)famous Minerva Press, which had a reputation for saturating the market with formulaic novels (Blakey, McLeod, Garside); these books published by William Lane took up between one third and one fourth of the market (Raven 73, Garside 83-84).¹³ Eight of the texts are epistolary, a type of fiction which averaged “at least 30 per cent of all novels published between 1770 and 1790” (Raven 31).¹⁴ Eight of the texts are anonymous. Three of the twenty-three contributing authors have two texts in the corpus.

Bibliographic details are given in the “[Primary Sources](#)” and in the [Chawton Corpus chart](#).

Although the Chawton novels were initially selected to be digitalized for their rarity, the assumption here is that they can be considered representative of domestic fiction written by women in the 1760-1830 period because of their high number and the sheer quantity of text they provide when grouped in a corpus.

2.2 The women corpus

This corpus adds novels by Eliza Haywood (*Betsy Thoughtless*), Frances Burney (*Evelina* and *Cecilia*), Maria Edgeworth (*Belinda* and *Helen*), Jane Austen (*Pride and Prejudice* and *Emma*) and Susan Ferrier (*Marriage*) to the Chawton corpus so as to give a perspective on differences and similarities between texts which have become canonical and texts which have not, bearing in mind the subjective dimension of “canon” and the serious changes it has undergone in the last decades for the period under study. Chawton House Library, its conferences, and the visiting fellowships it offers play a distinct role in the process of reevaluating writing by women. Some “canonical” texts were chosen so as to draw on earlier work I have published in the field of computer-aided literary analysis (e.g., [Betsy Thoughtless](#), [Evelina](#), and [Pride & Prejudice](#)).

The forty-two texts constituting the corpus WOMEN42 add up to about 5.8 million words, (i.e., roughly 50% more than the Chawton corpus). 41% of the Women corpus dates from before 1800, a proportion very close to that of the Chawton corpus. Bibliographic details are given in the “[Primary Sources](#)” and in the [Women corpus chart](#).

2.3 The reference corpus

Ideally, a Reference corpus is representative of the writing from the period and the genre on which the main corpora focuses. The availability of machine-readable texts limits this, but to a lesser extent than noted by Fischer-Starcke for her corpus established in 2004 (29).¹⁵

At this stage of the “Strategies of Writing” project, some degree of congruence with the other two corpora of the present study has been sought by not including overtly gothic or historical fiction: the texts selected had to answer the criteria of a narrative mainly set in the British Isles centered on a small group of characters. The same number of novels by women and men writers was chosen. This Reference corpus has been set up as close as possible in date and text size to the Chawton and the Women corpora so as to enable comparison and contrast (1748-1834). A larger reference corpus will be set up at a further stage of the project.

Clive Probyn’s *English Fiction of the 18th Century, 1700-1789* and Gary Kelly’s *English Fiction of the Romantic Period 1789-1830* served as guides to look for freely available electronic texts, mainly from [Project Gutenberg](#). Here again, some texts were chosen so as to draw on earlier work ([Roderick Random](#), [A Sentimental Journey](#)). The authors are Austen, Brunton, Burney, Disraeli, Edgeworth, Ferrier, Galt, Godwin, Goldsmith, Hays, Haywood, Hogg, Holcroft, Inchbald, Mackenzie, More, Opie, Peacock, Scott, Shelley, Smith, Smollett, Sterne, and Wollstonecraft. Ten of them have two texts in the corpus. Bibliographic details are given in the “[Primary Sources](#)” and in the [Reference Corpus chart](#).

Composed of thirty-four texts like the Chawton corpus and adding up to about 5.5 million words, the Reference corpus (CTROL34) partly overlaps with the Women corpus but not with the Chawton one. 41% of CTROL34 dates from before 1800 (the same as in WOMEN42) and 44% of the corpus belongs to fiction written by male authors (17 of the 34 works).

Other smaller corpora occasionally serve to confirm findings: they are made up of one author's fiction (Austen, Burney, Haywood) or composed of later works (15 texts of fiction 1860-1960).¹⁶

Now that the scope of the study and the content of the corpora have been made clear, computer-aided analysis will first be used in a global analysis of the Reference corpus. This analysis will establish that female texts do have specific traits (section 3) and from this, it will show how women writers use vocabulary (section 4). These findings will then be applied to the Chawton and Women corpora (section 5) and to one of the Chawton texts (section 6).

3. A common female vocabulary?

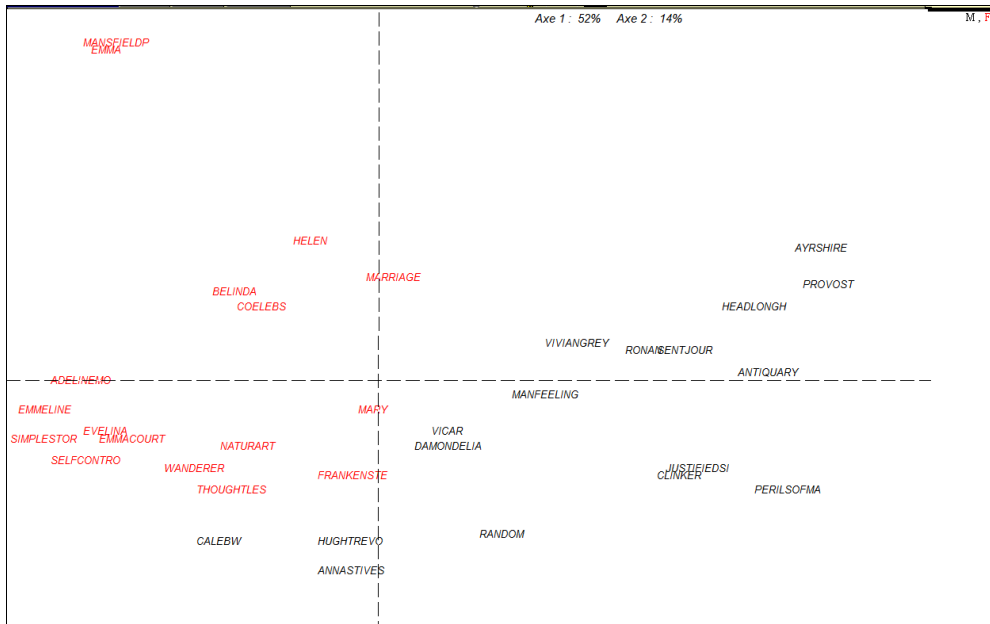
Exploration of the Reference corpus, made up of as many texts by women as by men, gives insight into features of fiction by women in the period and into “the ‘feminine’ novel” (the genre to which the selected Chawton novels pertain). The aim of this investigation is to tease out tendencies and “small touches” that can be viewed in isolation or grouped.

The tool used here is lexical connection, a mapping of the parts of a corpus (novels in the present case) according to the vocabulary they share (it consists in principal component analysis, provided by *Hyperbase*). A word contributes to drawing two texts together if it belongs to both, and a word contributes to pulling them apart if it only occurs in one of them (Brunet 60).¹⁷ Once this is assessed, principal component analysis makes it possible to plot the data onto maps which position the texts according to the vocabulary they share. This is done first by looking at the vocabulary, the presence or absence of words considered as types (section 3.1) then by taking into account the frequency with which the words are used (considered as tokens in section 3.2).

3.1 Words as types: The gender of writers

A clear distinction appears between texts by each gender when word-types serve to position the 34 novels in respect to one another (figure 1):¹⁸ this indicates that female texts share a common vocabulary. The study seeks to identify this vocabulary with the help of the three corpora.

Figure 1 – Lexical Connection on Types for the Reference corpus
(Principal component analysis, *Hyperbase*)
Titles in red by women, in black by men



Works by female authors (in red) are grouped in the left part of the map, and all but three texts by male authors are grouped on the right (Holcroft's *Anna St Ives* and *Hugh Trevor*, and Godwin's *Caleb Williams*).¹⁹ This venturing into female territory by the “philosophical” (Tompkins 296-311) or “political” (Spacks 222-53) novelists may serve to identify some of this stock of words (but this phenomenon is beyond the purview of the present study).

It is difficult, however, to identify the words which form this vocabulary and from which the map in figure 1 is drawn. After all, such graphic representation aims at synthesizing a huge mass of information, while output produced by software which can be analyzed in detail, such as lists of quantitative keywords, comes from the relative frequency of words, therefore considered as tokens (examined in the following section). The strength of quantitative analysis lies in the contrasts and comparisons it helps put in evidence in a corpus, more than in direct identification of what forms the common ground. Assessing what this common ground is, therefore, constitutes a challenge. Meaningful ways of assessing overuse or underuse of vocabulary in parts of the corpus against use in the corpus as a whole, or against another corpus must therefore be thought out. The role of a corpus is to provide a norm.

Computational stylistics (and software such as *Wordsmith* or *Hyperbase*) rely on the assumption that words are used with the same frequency throughout a corpus, which is valid as a null hypothesis with a large quantity of text. This makes it possible to contrast actual frequencies of word use with those expected (in virtue of the proportion of each part of the corpus). For this study, the corpora are divided into individual novels, which can be grouped into larger units so as to contrast features. Figure 1 indicates that it is useful to contrast the vocabulary of texts written by women and those written by men in the Reference corpus, and to check this against the female corpora. This will be done in section 4.

3.2 Words as tokens: The gender of characters

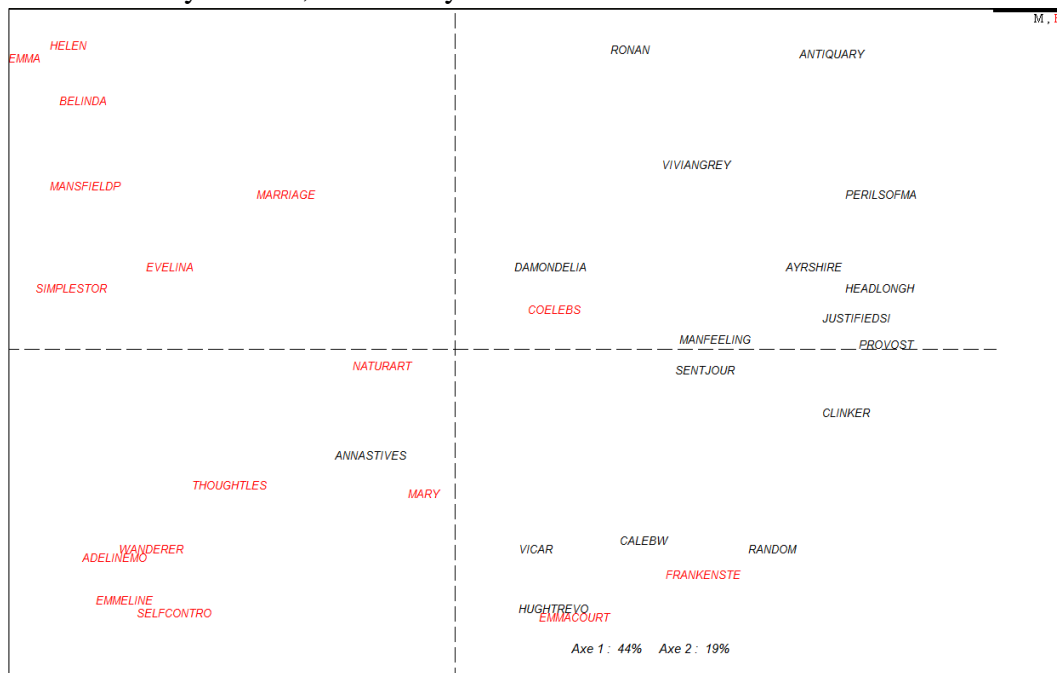
When frequency is taken into account, the mapping changes somewhat. Lexical connection on

tokens in figure 2 shows that this separates texts according to whether they focus on female or male characters: *Hugh Trevor* and *Caleb Williams* as well as *Coelebs* and *Frankenstein* have joined the “male” group on the right, while *Anna St Ives* joins the “female” one on the left. The “small touches” and tendencies drawn from this investigation in the following section are therefore expected to also be present in the Chawton corpus, made up of narratives written by women and for the most part focusing on female characters.

Interestingly, the strong homogeneity in vocabulary and frequency of use appears in all but one of the texts that both belong to the Reference corpus and form the canonical part of the Women corpus. Notice that *Emma*, *Helen*, *Belinda*, *Evelina* and *Marriage* are grouped in the same half of [figure 1](#) and in the same quadrant of [figure 2](#).²⁰ It will consequently be relevant to consider their collective features in order to identify those of “the ‘feminine’ novel.”

Figure 2– Lexical Connection on Tokens for the Reference corpus
(Principal component analysis, *Hyperbase*)

Titles in red by women, in black by men



A point must be made about the choice of Austen’s novels, whose homogeneity has been demonstrated repeatedly. Fischer-Starcke illustrates this homogeneity by testing the corpus of the six novels against one of “contemporary literature.” This corroborates Burrows (*Computation into Criticism* and “Style”), who distinguishes samples of writing by Austen from those by other authors, and Graves (“Vocabulary Profiles”), who “shows that the correlation in word usage is [...] stronger for Austen and Burney than it is for Edgeworth.”²¹ In Jockers’s corpus of 250 nineteenth-century novels, Austen “has one of the most consistent and unvaried styles” or, as he “put[s it] rather too bluntly, neither Austen’s stylistic nor her thematic range is exceedingly vast” (Jockers 93, 160-61). As three Austen novels only were needed for the present study, it was decided not to consider her Juvenilia and to discard the two novels published posthumously, *Northanger Abbey* and *Persuasion*. *Sense and Sensibility* was not retained

because it has two protagonists. *Emma* was selected as it centers on one heroine, whose name moreover echoes the “Julias and Louisas” made emblematic of novels in *Northanger Abbey* (I.14), a tendency confirmed by the high frequency of “heroines who lack a last name” among British novel titles from 1740 to 1850 (Moretti, *Distant Reading* 197). *Mansfield Park*, with its community wherein the heroine must find her place, and *Pride and Prejudice* as being midway between these two, were also chosen. *Emma*, present in both corpora, serves to ensure congruence, and acts as a reminder of the great closeness of the six novels.

Now that the presence of a vocabulary common to the female writers of the Reference corpus has been established, other statistical tools can be used to identify it. Most of the women authors of the corpus wrote novels centered on female characters. Their specific vocabulary can therefore be extracted by taking their frequency into account (words as tokens).

4. Assessing female use of vocabulary

One way to isolate the vocabulary that draws texts together is through “keyness” or “specificity,” the statistical identification of quantitative keywords. These are “words which are statistically salient in a text or corpus compared to a reference corpus” (Fischer-Starcke 68), and these terms are used proportionately more (or less) frequently in one part of the corpus than a regular distribution over the whole would warrant (the null hypothesis mentioned earlier).²² As Fischer-Starcke states, “keywords function as pointers to literary meanings in a text” (69).²³

The list of words that are most specific to texts written by women grouped into a sub-part of the Reference corpus confirms the hypothesis which comes to mind: female pronouns and forms of address head the list when proper names are discarded ([figure 3](#)). Indeed 82% of female pronouns are found in the texts written by women, which make up only 56% of the Reference corpus:²⁴ the imbalance is pronounced and the proportion remains the same if terms of address are added on. This proportion means that 28 occurrences of female pronouns are used for every 1,000 words in female-authored texts but only 8 for every 1,000 words in male-authored texts. This is, of course, content-related: *Frankenstein* has 7 female pronouns for every 1,000 words. The rate is identical in the Chawton corpus. However, as texts by women in the 1860-1960 corpus also use on average 28 female pronouns for every 1,000 words, this does not seem limited to the texts chosen for the present study.²⁵

As for male pronouns, women writers prefer men as subjects and objects rather than as possessors (grammatically speaking: they prefer *he*, *him* rather than *his*, which is the top pronoun in the negative specificity list for the female part of the Reference corpus, [figure 4](#)) and their texts are highly concerned with male characters, using 52% of male pronouns while making up 56% of the Reference corpus (on average, 21 male pronouns for every 1,000 words versus 23 in the male part of the corpus). The difference becomes stronger with the many terms used for males in the “negative specificity” list (i.e., words scarcer than average: *master*, *man*, *men*, *fellow*, *gentleman*, *captain*, *son*, and *boy*). When the occurrences of these terms are added to the male pronouns, the proportion of frequent terms referring to males is brought down to 50% in the female part of the corpus, significantly less than the expected 56% seen in a regular distribution over the entire corpus.²⁶

When words used more in the female texts of the Reference corpus are examined in detail, patterns emerge: among the 266 terms which are not names, 35 can be grouped as pertaining to feelings, 13 to family and marriage, 9 to time, 7 to sight, and 7 to speech.²⁷

I focus here on terms that express feelings or emotions: lemmatizing them (adding the inflected or variant forms of the same word) leads to a list (labeled “EMOTIONS”) of 108 items, which can then be used with any corpus as a litmus test to assess whether this topic is important enough to warrant more detailed investigation.²⁸ This paves the way for studies combining quantitative stylistics to approaches like that of Thomas Dixon’s analysis of the meaning of tears over particular periods (“Enthusiasm Delineated” and *Weeping Britannia*), or to the manner in which feelings and emotions were described in physiognomy, as the popularity of Lavater’s 1775-78 *Physiognomische Fragmente* culminated at the beginning of the nineteenth century and clearly influenced some of the works in the corpora (this will be developed in my larger “Strategies of Writing” project).

Globally, 12 items from the EMOTIONS list are used for every 1,000 words in the Women corpus, and only 6 for every 1,000 words in the male part of the Reference corpus (this is the same in the 1860-1960 corpus as a whole, with hardly any distinction between texts by male or female authors). The Chawton corpus has a marginally higher use than the Women one.²⁹ Quite logically, 70% of EMOTIONS belong to the female part of the Reference corpus, but of course with variations: these words are used frequently in *Frankenstein*, also by Austen (*Mansfield Park* and *Emma*) and Edgeworth (*Belinda* and *Helen*), but not so frequently by Burney (*Evelina*, *The Wanderer*, [figure 5a](#)).

The Women corpus confirms Burney as a low user of EMOTIONS, with *Cecilia* intensifying this tendency ([figure 5b](#)). The double graph also shows that within a context of female fiction rather than compared with the Reference corpus, Austen becomes an average user, which proves the appositeness of inserting texts in different corpora. As for the seven Minerva novels, the diversity Blakey and McLeod note in the output of William Lane’s press is illustrated here: their use of EMOTIONS terms is not consistent, contrary to the clichés about them.

The analysis of specific vocabulary or quantitative keywords has brought to light several characteristics pertaining to “the ‘feminine’ novel” written in the period covered by the Chawton novels. Some of them are “subtle ‘features’” such as pronouns (Jockers 64) and others consist of thematic or topical groupings. These common characteristics include:

- an interest in both genders, whereas fiction by men in the control corpus takes females into account to a much lesser extent;
- a strong use of small-group interaction and of dialogue, with a particular liking for *cried* which points to the prevalence of intensity;
- a concern for feelings and emotions, along with (not detailed here) family and marriage, and sight.³⁰

The following section will carry out the same type of analysis on the Chawton corpus.

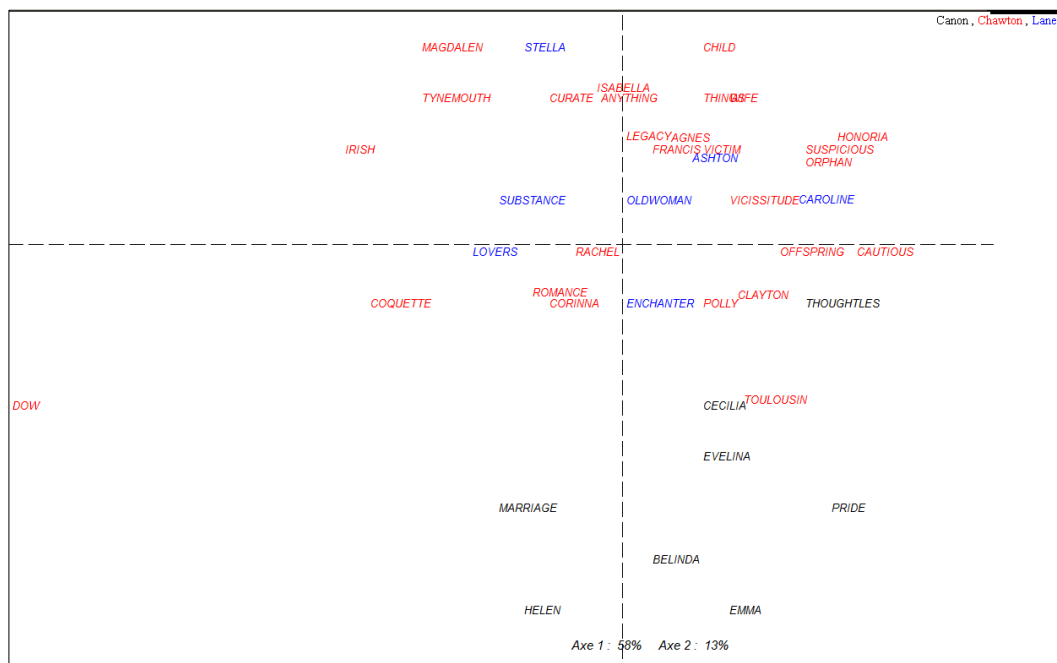
5. The Chawton Novels

What follows considers the Chawton novels both on their own and in the larger Women corpus, with a focus on individual texts or groups of texts. As noted earlier, the methods used in this study make it easier to spot the causes of difference than the causes of normality.

5.1 General view

The mapping of the texts displays typicality and eccentricity. An example of a text that is mapped as highly eccentric is *Three Weeks in the Downs*: when lexical connection is computed from word-types (the vocabulary), this text is positioned far away from all the other texts (DOW in [figure 6](#)) but this is not the case when it is computed from tokens ([figure 7](#)). The list of the text's disproportionately frequent terms compared to the Chawton corpus (positive specificities) shows that most of them have to do with seafaring.³¹ Moreover, *Three Weeks* ranks first for the use of words unique in the Chawton corpus (*hapax legomena*) and second after Ferrier's *Marriage* in the Women corpus, which indicates a greater diversity in its vocabulary than the other novels. The rest of the title gives away the reason for its presence in the corpus: *or Conjugal Fidelity Rewarded / Exemplified in the Narrative of Helen and Edmund / Founded on Fact / By an Officer's Widow (Three Weeks)*. Its originality comes only from its setting, not the way in which it uses vocabulary.

Figure 6 – Lexical Connection on Types in the Women corpus (*Hyperbase*)
Canonical novels in black, Minerva novels in blue, non-Minerva Chawton novels in red



The seven Minerva texts are positioned among the rest of the corpus. They appear in two groups for types ([figure 6](#) in blue) but are dispersed for their use of tokens ([figure 7](#)). This provides another indication of their diversity, and shows that they blended with the popular literature of their time. Because the Lane texts are not grouped in one part of the lexical connection graphs as

were those written by women or telling the stories of female characters in the Reference corpus, it does not make sense to contrast their vocabulary as a whole to that of the rest of the corpus. Small clusters appear, however. *Ashton Priory* is very close (in terms of types) to another Lane text, *The Enchantress*, and close to *The Old Woman* and *Caroline*, all published within a fourteen-year time span, 1787-1801. To some extent, this suggests a fairly standard Minerva Press vocabulary, but as explained earlier, delineating this common stock of words is difficult. The dissimilarity in their use of tokens indicates that each text uses this vocabulary with its own frequencies, thereby producing an impression of variety. A large-scale project on a high number of machine-readable Minerva texts would certainly yield interesting results and make it possible to complement the analyses of Blakey and McLeod.

The opposite phenomenon of a greater similarity in the use of tokens than of types can be observed for texts by the same author, particularly Frances Jacson's *Isabella* and *Things by their Right Name* (THINGS, [figures 6](#) and [7](#)). A similar proximity was observed for Austen's texts in the Reference corpus. This gives validity to the hypothesis that authors have stylistic signatures, and this may help solve some attribution questions.³² The greater proximity of *The Corinna of England* (CORINNA) and *Substance and Shadow* (SUBSTANCE) in the larger Women corpus ([figure 6](#)) than in the Chawton corpus ([figure 8](#)) may reinforce the attribution of these two texts to Mrs E. M. Foster on the Chawton website, an attribution that is at odds with the skepticism of *British Fiction, 1800-1929*: "The attribution to Foster apparently hinges on a chain of attributions within titles, stretching to 1817, and must without further evidence be considered unlikely." Detailed work with a corpus including texts by the other potential authors would be needed to test this hypothesis. The 1810 *Corinna* may have gotten Foster into Minerva Press for the 1812 *Substance and Shadow*.

With *The Castle of Tynemouth* (TYNEMOUTH or TY) and *Any Thing but What You Expect* (ANYTHING), Jane Harvey seemed able to diversify her use of vocabulary somewhat, much like the canonical writers ([figure 7](#)), although her fiction was deemed "very harmless food for the circulating library" (*Flowers of Literature*, 1806, lxxx, quoted in *British Fiction, 1800-1929*). Here again, a close analysis would probably prove rewarding.

The eight epistolary novels are also close in their use of tokens but not in their use of types: letters imply repetition of particular terms but tell different stories. *The Suspicious Lovers*, *The Offspring of Fancy*, *The Victim of Fancy*, *Agnes De-Courci*, *The Cautious Lover*, *Vicissitudes in Genteel Life*, *The Old Woman*, *The Wife* form a cluster in [figure 7](#). This is confirmed by the Women corpus where Burney's *Evelina* belongs to the same quadrant ([figure 8](#)). As can be expected, they are characterized as a group by a strong presence of first- and second-person pronouns (respectively 33% and 22% while making up only 17% of the Women corpus). These markers of direct address between letter-writer and correspondent unsurprisingly combine with an overflow of *soul* (34%) and *heart* (25%). However, the EMOTIONS list as a whole is not overrepresented (19%). The novel which is closest to them in the Chawton corpus ([figure 8](#)) is entitled *Memoirs of Count Touloussin, Written by Himself*: the first person duly presides, its subjectivity foregrounded by the main title, *Prepossession*. It is probable that more features of epistolary fiction could be teased out from a systematic study.

It has been possible so far to examine several texts more closely, and one in great detail.

5.2 Focus on individual texts

The anonymous *Ashton Priory* is mapped centrally because of the frequency with which it uses vocabulary (tokens) in the Chawton corpus ([figure 8](#)). With its eponymous “most extraordinary pile,” the Priory that results from “such repairs and supplementary erections as suited either [the successive proprietors’] choice or convenience” (Chapter XVI), *Ashton Priory* is not unlike *Northanger Abbey* in producing expectations of a gothic novel but thwarting them. As a 2011 blogger regrets, this seems to work on twenty-first century readers as it probably did on quite a few circulating library predecessors.³³ It also resembles *Cecilia* with its rich heroine moving from one guardian’s house to another. Revolving around catastrophic plans devised by the older generation for the younger one, it provides a more Manichean ending than Burney’s 1782 text: while the deluded victim of “chimerical ideas” acquired from the “trash” she has read, “far gone in the romantic taste,” dies (Ch. LI, Ch. XIII), the heroine is happily married, having regained her lost “fortune” after many “incidents.” The terms within quotation marks in the preceding sentence appear among the few quantitative keywords, both in the Women corpus and against the Reference one, along with “occasion,” “affair,” “merit,” “sentiment,” and ... “sum”: in *Ashton Priory* one finds a lively and sordid tale of sex and money with a variety of settings (London, Germany and Somerset). Because of its central position for the use of tokens, it can be regarded as a template from which others deviate. Despite some self-deprecating irony, notably in chapter headings, it also seems to justify Austen’s call for solidarity among novelists, as the heroines definitely do not “patronize” heroines of other novels: *Ashton Priory* rather goes for novel bashing.³⁴

Fairly central for its use of tokens, *The Unexpected Legacy* (*LEGACY*) comes from the pen of Rachel Hunter, whose voluminous lachrymose fictions Jane Austen mocked in an 1812 letter to her niece Anna: “Miss Jane Austen’s tears have flowed over each sweet sketch in such a way as would do Mrs Hunter’s heart good to see; if Mrs Hunter could understand all Miss Jane Austen’s interest in the subject she would certainly have the kindness to publish at least 4 vols more about the Flint family,” a reference to the 1806 *Lady MacLairn, the Victim of Villainy* (*Jane Austen’s Letters* 203). Tears duly overflow in the 1804 *Unexpected Legacy*, but only about 60% more than the average in the Women corpus. *Agnes de-Courci* sheds more than twice the expected quantity of tears, while Austen comes out driest with only a quarter of the Chawton average and even fewer as she got older, from 23 occurrences in *Sense and Sensibility* to only one in *Persuasion*. The word *tears* belongs to the list of key terms in the female part of the Reference corpus from which the EMOTIONS list was established, but as [figure 6](#) shows, *The Unexpected Legacy* is not a high user of the list as a whole, and in this novel at least, Hunter prefers smiles to tears, with more than twice the expected quantity of the former (lemmatized), of which she is the highest user in the Chawton and Women corpora (*Lovers and Friends* and *Helen* are not far behind). Detailed work on the expression of emotions and the use of terms pertaining to the body is underway.

Despite Austen’s bantering, *Emma* and *The Unexpected Legacy* are not far apart for the use of types (according to the first factor, which positions texts along the horizontal axis and carries much more weight than the second, which separates them along the vertical axis, [figure 6](#)).³⁵ The use of tokens pulls them apart, clustering all canonical texts except *Betsy Thoughtless* and *Marriage* at a safe distance from most Chawton novels ([figure 7](#)).³⁶ From this, it can be inferred

that Austen, Burney, Edgeworth, Ferrier, and Haywood drew from the same stock of words as did their less famous female contemporaries, but they used them at different rates. The difference between these canonical texts and the rest of the Reference corpus established earlier ([figure 2](#)) provides a basis from which to identify common vocabulary traits. This approach makes the less famous fiction a benchmark, rather than providing a normative judgment of texts that left a small footprint in literary history.

One may also turn to satire and conscious exaggeration to gain insight into the ways a genre was perceived by contemporaries. Sarah Green's *Romance Readers and Romance Writers: A Satirical Novel*³⁷ uses word types not very differently from the most central texts in the Chawton and Women corpora ([figures 7](#) and [6](#)) and this satire keeps company with non-canonical texts when frequencies are taken into account ([figure 7](#)). Like William Beckford's burlesques *Modern Novel Writing, or the Elegant Enthusiast* and *Azemias* (1797), *Romance Readers and Romance Writers* "holds up a mirror to a generic landscape," partly by blending into it in terms of vocabulary and word frequency (Gemmett 20). A comparison with Green's other fiction such as *Gretna Green Weddings* or *Scotch Novel Reading* would help define her strategies to manage different degrees of satire.

Corpus analysis can therefore be useful in the decision of which texts to digitize. It also helps to identify texts worth analyzing in detail so as to gain insight into a genre. *Ashton Priory* will be the object of a publication to come. Jane Taylor's 1817 *Rachel*, the most central Chawton novel for types in the Women corpus ([figure 6](#)) and fairly central for tokens ([figures 7](#) and [8](#)) is examined as a case study in the following part of the present essay.

6. Rachel: A "Highly Original" tale told in unoriginal terms

The Orlando Project describes Jane Taylor's anonymously published 1817 text as "the highly original *Rachel. A Tale*, with an ugly heroine and an anti-romantically named male protagonist, Tomkins, [... who] never concludes his courtship" (Brown, Clements, and Grundy).³⁸ Yet *Rachel*'s central position in the lexical connection graphs indicates that it was written with the most common vocabulary used in unexceptional frequencies: in this respect, it exemplifies the norm of the corpus.³⁹ The seeming paradox of a highly original tale told in unoriginal terms is examined by focusing on words or groups of words foregrounded by the software or found striking on reading the text if this perception is confirmed by usage in the corpus.

One hypothesis for *Rachel*'s centrality could be its size: with some 34,000 words (tokens) it is the second shortest text in the corpus. Yet *The Reward of Virtue; or the History of Miss Polly Graham*, the shortest one with some 33,000 tokens, is positioned away from the center of the graphs for measures of lexical connection both in terms of vocabulary and of frequency in the Chawton and Women corpora ([figures 6](#) to [8](#)).⁴⁰ *Polly* consists in the independent misadventures of deserving characters, mostly but not only female, loosely linked to "the worthy community at Bounty-Hall," an "asylum" for "virtuous women of family [...] reduced, through unavoidable misfortunes, to poverty" (chapter IX). The Type Token Ratios of *Polly* and *Rachel* are very close (0.11) and the two texts have a comparable number of *hapax legomena*, words unique in the corpus (respectively 63 and 56 in the Women corpus). The size of *Rachel* is therefore clearly not the determining factor for its unexceptional vocabulary.

A more plausible hypothesis is the text's didactic purpose, which also underlies its more famous predecessor *Display* and the 1821 *Prudence and Principle*, both subtitled, like Rachel, "A Tale" (but not available in electronic form). Peter Garside distinguishes Taylor along with Maria Edgeworth and Rachel Hunter for writing books that target young audiences while "evidently enjoy[ing] a wider currency" and having "effectively entered the mainstream," contrary to other "fiction for juveniles" excluded from the lists of *The English Novel 1770-1829* (35). Didacticism runs through the corpus in varying degrees, often to be taken *cum grano salis* in texts openly written for entertainment, but perhaps most seriously of all in Hannah More's 1809 best-seller *Cælebs in Search of a Wife* (Demers 8), included in the Reference corpus.⁴¹ As will be seen, More's influence on Taylor is suggested by several parallels in the use of vocabulary. With a narrower scope, Taylor's *Rachel* has a well-structured and focused story with a simple and common vocabulary.

Rachel has most of the features which characterize female texts as identified in the Reference corpus (section 4), hence its typicality. The interest in both genders appears clearly, with the use of female pronouns slightly higher than average (32 per 1,000 words), and with male pronouns slightly lower (18 per 1,000 words). The frequency of terms for the family and marriage, as well as for time and sight does not differ from the average in the Women corpus. Emotions are somewhat less frequent. As shown below, they are replaced by another topic, which also explains that dialogue verbs are more frequent than average. Small-group interaction plays a particularly important role in the didactic dimension of the text, creating a fairly safe environment in which the heroine can develop through trials and errors.

In order to make Taylor's writing strategy apparent, several salient terms from *Rachel* are analyzed in comparison with their use in the three corpora.

6.1 A different use of "Foil"

Like most other texts of the corpus, *Rachel* tells a story of courtship. It involves two main young women, three secondary ones, and one potential husband, Tomkins, who is introduced in the tale through a response to an invitation: "I will go, and who knows (looking round his solitary apartment with a kind of half sigh) but I may find a wife amongst them." The idea pleased him, for he was thirty, and had long been desirous of "taking to his bosom a wife" (Taylor 1). "A man of sober habits," he accompanies his guests to church on the day following his arrival and is placed beside three women as the family pew cannot accommodate him (3). With such a configuration of characters, Taylor combines the objectivity of a third-person narrator with a male focalizer in the description of his neighbors:

The eldest was a middle-aged woman, in whose face might be read marks of a strong understanding, mixed with great reserve [...]. The second, though indeed the first that attracted the notice of our visitor, was much younger: to a tall genteel figure, she added a remarkably interesting countenance; and as Tomkins gazed on the pensive expression of her eyes, he thought he had at last found the wife he had so long been seeking. The third, and last of the party, appeared as if intended for a foil to her companion: she was tall, stout, and awkward in her appearance; while

her face, devoid of all pretensions to beauty, expressed nothing to excite either interest or admiration. (4)

When Tomkins formulates his impressions for his host, this description becomes “one is very proud [which tallies with public opinion], another very pretty, and the third most incorrigibly stupid” (6). Somewhat unexpectedly, his interlocutor attributes the last adjective to the name put forward by the title: “‘Alas, poor Rachel!’ cried Mr. Simpson, laughing” (7). Before examining by comparison with the rest of the corpus whether the use of the adjectives applied to the eponymous heroine is as noteworthy as the surprise effect created by Taylor suggests, a look at “foil” in the quotation above gives insight into the way Taylor’s rather subtle didacticism works.

Notice how Taylor reinforces “appeared” with “as if intended” when introducing “a foil to her companion” in Tomkins’s observation that Rachel “appeared as if intended for a foil to her companion.” With this word choice, the reader can hardly escape Taylor’s hint that appearances will be proven false. Indeed, “appear” in its different forms and derivatives is used twice as frequently in *Rachel* as in the Women corpus norm.⁴² Comparison with the Reference corpus suggests that Taylor uses her single “foil” in accordance with the conventions of contemporary female writing. In the Women corpus, only 7 of the 30 occurrences of the term are used for persons and 4 for a comparison of women.⁴³ Of these, two come through a first-person point of view and two through a third-person one. Like *Rachel*, *The Enchantress* introduces “foil” among the words of a third-person narrator describing a situation which suggests a male focalizer:

Mrs. Macfarlane had seated Sir Philip next herself on a sofa; and it was easy to see that he needed to take but very little pains to obtain the best graces of his captivating hostess. The women of the party were not attractive. Some old maids and dowagers served as foils to Mrs. Macfarlane; who, though her countenance was not engaging, had fine eyes, and a very clear and animated complexion. (Martin 9, emphasis mine)

One difference with *Rachel* resides in the earlier description of the lady, “having a physiognomy as crooked as her shape” and being labelled “the crooked lady” or “the crooked step-mother” 10 times in the text (*The Enchantress* uses 11 of the 31 occurrences of “crooked”). Taylor does not make Rachel deformed, only “tall, stout, and awkward in her appearance,” leaving clear room for improvement of at least the third characteristic.

The other two texts of the corpus that use “foil” for women, both epistolary novels, integrate the term within a narrower first-person point of view. The word serves to express a male judgment in *The Cautious Lover*:

a Mrs. Staples, one of your *good pretty kind of women*.—She is tall and slim; she has a pale complexion, black eyes, but no chin.—She smiles and simpers at every thing that is said.—I am amazed that a woman of Charlotte’s understanding can like such a non-entity to be always near her; as I know she is above all that meanness of carrying a foil about with her;—she has indeed no sort of occasion for one, her beauty wants nothing to set it off. (Letter XLIII)

Charlotte's justification for the presence of such a character is what "sets [her] off" for the reader, if not for the male letter-writer. At a point in the narrative where suspense is important, the end of the first volume,⁴⁴ Charlotte explains her generosity by stating that Mrs. Staples is "the daughter of a particular friend of my mother's [...] who was married unfortunately," as the letter-writer reports to his male correspondent (Ibid.). Contrary to *Rachel*, *Cautious Lover* does not apply "foil" to the main character, but to one he deems "unable to raise envy in woman, or desire in man." In *The Wife*, the point of view is female and self-applied, but negatively so: "Surely I shall not become a foil to this Miss Sommerton," the heroine's friend writes to her, seized by "a fit of humility" after stating, "I really was a very striking figure" while relating her "first appearance" at Bath (Letter II).⁴⁵

Rachel tells of a less openly aggressive rivalry between the two main characters and their three young female neighbors, giving the use of "foil" to the third-person narrator only, so that the reader will, like the hero and the heroine, learn to go beyond appearances and later revise this assessment. This process develops at least in part by overcoming the negative connotations of the adjectives used when introducing Rachel into the narrative, which will now be examined.

6.2 Awkward, tall and stout

Rachel declares one page before the end "I have learned a lesson that will, I hope, be of use to me all my life" (152). This lesson concerns her behavior rather than her physical appearance as she has put herself in a compromising position by shielding her wayward brother. However, a few pages earlier, she is made to reflect on the reformable aspect of her person by appropriating one of the three adjectives initially applied to her by the narrator: "I begin to perceive, that a total want of refinement is almost as bad as a superfluity: in other words, I would say, that *my awkwardness* is becoming troublesome to myself, and I am, therefore, desirous of shaking it off" (140-41, my emphasis). The association of *my* to *awkwardness* is rare, even unique in the Women corpus, but Austen likes the word and its derivatives: *awkward*- appears 3 to 17 times in each of her novels. Like Rachel, Fanny Price in *Mansfield Park* is first described with the adjective by the narrator, and later uses the noun self-reflexively.⁴⁶ This may come from *Coelebs in Search of a Wife*, the only text in the Reference corpus which has the collocation.⁴⁷

Possibly taking her cue from Austen and most probably from More, Taylor uses "awkward" and its derivatives significantly more than the average of the Women corpus. *Rachel* comes even before *Emma* (the highest user in the Reference corpus), with 8 occurrences rather than the 1 to 2 expected, mostly for the heroine.⁴⁸ After the initial "tall, stout and awkward in her appearance," the forms of the word are spread out throughout the text. The gradation suggested by "her manners were more than usually awkward when the divine was present" (25) leads to explanatory reformulations: "the natural reserve, or, as it may perhaps more properly be termed, the awkwardness of Rachel" in the second chapter (48) and "the awkwardness, arising from the seclusion in which she had been educated" in the third (57). This serves to stress that she possesses other qualities: "Where the exertion of her fortitude was required, it had seldom been known to fail" (57). Her propensity to action, however, puts her in trouble as she is seen with a young man whom nobody knows is her brother. Yet, she learns from this trouble, to the extent that she finally appropriates the discourse of the narrator and refers to "my awkwardness" in the passage quoted earlier (141). Taylor links awkwardness to the family as the two other

occurrences of the term appear in dialogue spoken by Rachel's wayward brother, who applies it to situations rather than persons.⁴⁹ The only other occurrence of the term qualifies an intervention by the reader, of which more later.

Of the other two adjectives initially attached to Rachel by the narrator, "tall" is neutral, both in the text and in the Women corpus. The description of her companion Sophia in terms of "a remarkably interesting countenance" joined to "a tall genteel figure" (4) shows that size is not an issue for Taylor. In the Women corpus, "tall" applies quite expectedly to almost twice as many men as women and participates in nearly as many positive as negative characterizations (respectively 24 and 27 occurrences).⁵⁰ "Stout" on the other hand, when applied to women (7 of the 36 occurrences in this corpus), has connotations of a low social status: it serves for a nurse (*Belinda*), a servant-girl (*The Child of Mystery*), a threatening gipsy (*Emma*), and the landlady of an inn (*Three Weeks in the Downs*). But the undesirable fate of Lydia Bennett, "a stout, well-grown girl of fifteen" (Austen, *Pride* 90) does not endow the word with the positive connotations of childhood (2 occurrences) when applied to middle-class heroines. As Rachel's father is "a respectable haberdasher in London" (56) which makes the family belong to this middle-class, whether the adjective refers to strength or corpulence must remain conjecture.⁵¹ Both point to a lack of delicacy and refinement, the qualities possessed by her cousin to whom she seems to serve as a foil.

Far more damaging, as we shall see, is the accusation of "idiocy" and "stupidity."⁵²

6. Intelligence?

Taylor's hero, Tomkins, refers to her heroine as, "The third most incorrigibly stupid" (6). It takes the entire text for Tomkins to become certain that he "had found an intelligent being, with whom he would have been well contented to pass the remainder of his life" (153), although Taylor does not make him conclude his suit. The vicar who increases Rachel's awkwardness also deems "her intellect [...] weak" before launching "on his favourite hobby-horse" (36), an enlightened plan of education for girls.⁵³ The vicar finally comes round as well: "But even he now declared, that he had mistaken Miss Porter's character; and that it required a mind somewhat stronger than he had imagined her to possess, to have acted with so much resolution" (149). Tomkins really wonders about the intellectual powers of Rachel in the middle of the narrative, while fittingly helping her to climb over a gate which cannot be opened (80). This symbolic act comes in between two close occurrences of "after all she must be an idiot" (80, 83), which are framed by the other two uses of the noun.⁵⁴ The last use confirms Tomkins's better appreciation of Rachel when she answers his indirect accusation of her by a reflection on ends and means: "he could hardly credit his senses, that it was delivered by a girl he had so recently pronounced an idiot" (85). The first, much earlier use, serves as a way for the narrator to discredit the general perception of her: "In all company she was so remarkably silent, that her acquaintance had long since agreed, that she could be but little short of an idiot; for as a woman's inclination to talk was never disputed in any age, a far more fallacious conclusion might have been drawn, than that if she was silent it was from want of ability" (25). Because Rachel does not generally engage in conversation, she is tentatively destined to a man who matches the first adjective allotted to her: he is "universally noted for his stupidity" (24). He will be discarded precisely because of this feature, fighting rather than getting the ladies out of harm's way in a crisis (123) while she, by that point in the

narrative, has displayed good sense and practicality: taking care of an injured servant after her cousin has refused to do so (99, 110), mending a badly torn dress during a walk (118), and most of all, managing to shield her brother at the expense of her reputation. In Tomkins's interior debates about her which the narrator makes the reader share, her secret involvement with the unknown young man is a clinching point: "He had been deceived in Rachel, in every respect. At first he had thought her a fool; afterwards he had imagined that amiable qualities were concealed beneath her apparent stupidity; but he had now learnt that she had wit enough to carry on an intrigue" (82-83). "Wit" has been used for her before, but rendered negative by the unworthiness of those from whom this judgment emanates and by the immediate context of the word: "her parents (like most others) beholding in their offspring qualifications that would have escaped the observation of the most accurate observer, had persuaded themselves into a belief that their Rachel was a wit," because she "ridicule[s] the conduct and behaviour of her companions" (47). The possessive ("their Rachel") and the verbal construction ("was a") reduce her to this characteristic, whereas Tomkins gives her the credit of being endowed with the possibility of using her mind ("had wit enough to"). If the two uses of "wit" prove that she is intelligent, they do not put her intellectual powers in a favorable light: she seems to be able to use them only for slander or intrigue. After Tomkins's recurring use of "idiot" combined with "her apparent stupidity," the second half of the text establishes her worthiness and her realization that she should discard the awkwardness that serves her ill.

With the guidance of the narrator, the reader sees through the "incomprehensible Rachel" (87) in a way Tomkins cannot, in part because of his "prejudice [...] against learned ladies," which introduces the theme of idiocy early on: "he would have found it hard to say, whether the open-mouthed idiotic stare of wonder disgusted him more than a lecture on causes and effects, pronounced in the most elegant language by a more enlightened female" (6). He doubts this prejudice when he discovers that the book dropped by the aptly named Sophia is a Latin testament (5), another link with *Cælebs*.⁵⁵ However he then prudently qualifies the girl with "a remarkably interesting countenance" (4) merely in terms of her appearance and with a determiner connoting vagueness rather than giving explicit ranking: "one is very proud, another very pretty, and the third most incorrigibly stupid" (6). As the narrative develops, "the intelligent and beautiful Sophia" (81) loses her appeal when she is found engrossed in another book, "the most nonsensical composition chance had ever thrown his way," the title of which we are spared (98). She moreover sees her cousin as "the coarse, vulgar, illiterate Rachel" (a cruder version of the narrator's initial "tall, stout and awkward"), while both Tomkins and the reader by then know otherwise (90).

Rachel is one of the two texts to stand out for the frequency of "idiot" and its derivatives in the Women corpus (1 occurrence at most is expected in a text the size of *Rachel*, as there are 55 in the corpus). In the other, *Lovers and Friends* again,⁵⁶ these terms are scattered, used 12 times by and for five different characters, whereas 4 of the 5 occurrences in *Rachel* apply to the hero's view of the heroine. Taylor further focuses her narration by not having recourse to the weaker, four times more frequent "stupid" and its derivatives in a quantitatively noteworthy way.⁵⁷ To question intelligence, she prefers to repeat "intellect-", with 4 occurrences rather than the 1 expected.⁵⁸ This reveals more subtlety than in Jacson's *Isabella* which dwarfs all other Chawton novels for the use of this lemma (21 occurrences rather than the 3 expected) by including "intellect" as one of many constituents which make characters' mechanistic bodies react to

stimuli (“the intellect of Isabella was bewildered,” Ch. XXXII), or as the eponymous heroine is told when encouraged to manage without the fortune her husband has squandered but within the threesome he imposes, “God has given you an intellect to comprehend, an integrity to support the difficulties that surround you” (*Isabella*, Chap XIX).⁵⁹ In *Rachel* intellect is not presented as a tool at the disposal of characters but it belongs to male discourse on female intelligence, either the discussion between Tomkins and the vicar, or Tomkins’s improved perception of Rachel given through internal focalization three quarters of the way through the text:

Still he was undetermined whether to give her credit for a cultivated understanding, as she always appeared to shun all conversation which required much intellect or learning to support it; yet he had observed that she paid the most profound attention to it, when carried on by others. More than once he had seen her lips part, as if she were about to declare her sentiments; but the inclination was always repressed, and a slight blush usually succeeded the involuntary motion. (112-13)

Rachel features among the texts which use the forms of “understand” more than average in the Women corpus but it does not stand out for it. Rather, Taylor relies on focused but not insistent repetition (4 to 5 occurrences) of two fairly rare, but not unusual, terms and their derivatives, “idiot” and “intellect,” combined with words also frequent elsewhere to make the questioning of intelligence one of the dominant topics of her tale. *Rachel* comes second behind *Lovers and Friends* and before *Isabella* in the Chawton corpus, and fourth in the Women one for its use of terms expressing intelligence, after “clever” *Emma*, Edegeworth’s *Belinda* and *Helen*, and equal to *Marriage* and *Lovers* ([figure 9](#)).⁶⁰ The debate on intelligence provides Taylor with yet another means to teach the reader the importance of going beyond appearances.

6.4 Questioning

The process of questioning is also a dominant topic in Taylor’s narrative. As has been pointed out, Taylor’s tale largely revolves around Tomkins’s observations about Rachel’s intelligence. More generally, terms related to interrogation feature among the vocabulary specific to *Rachel* in the Women corpus: “desirous,” “inquire,” “opinion,” “demanded,” “asked,” to which must be added more than double the number of question marks than the norm in this corpus. Only Burney’s *Evelina* and *Cecilia*, and the anonymous 1771 *The History of Lord Clayton* rank above *Rachel* for interrogative markers.

The questioning process plays an important part of Taylor’s didacticism, particularly as it involves the reader, clearly a young female (“my fair reader,” 1). She is expected to understand the purport of the absence of closure better than the character to whom the last word (or rather paragraph) is given: “As, however, he has made no one his confidant, time alone can discover in whose possession he left his heart; but it is expected his next visit to E— will ascertain the fact” is followed by “Mrs. Jenkins, even now, whispers with a significant nod, and no less significant glance, that that period will prove which of her daughters is to be the wife of Mr. Tomkins” (Taylor 153). All the Jenkins women have been discredited: the mother is ridiculed for her garrulity (9-16) while Tomkins deems Rachel’s taciturnity only excessive (113), the eldest daughter is dismissed early on by him as “one of those commonplace beings, known by the name

of young ladies” (28) although she plays a significant role in the story; the second is disqualified by Rachel (“I am almost tired to death with an account of Mary Jenkins’s dog’s toothache [sic]. I wonder how people can talk such nonsense!” 41), and the third is not even endowed with a name. The reader can therefore easily guess who Tomkins will come back for (the title of the book helps, of course).

Such subtle didacticism, intimated rather than stated outright, partly rests on the presence of the “reader” at the beginning of the tale, with a reminder in the middle.⁶¹ Both occurrences are linked to questioning, with the middle one simultaneously stating and denying agency: “‘After all what?’—is a question that requires a long answer; and whether the reader may be disposed to ask it, or not, it is almost absolutely necessary to give it, to account for Tomkins’s subsequent conduct” (80). The question refers to Tomkins’s “after all she must be an idiot” (80), reiterated three pages later, and ensures that the reader thinks along with the male character. The focus on the reader at this point of the text is reinforced by three occurrences of “may be” in close succession: “whether the reader may be disposed to ask it” is followed by the often encountered “It may be remembered”⁶² and the far less conventional “though [Tomkins] may be deemed a spiritless, droning, sanctified bore” (81), rather than the dashing hero a reader might wish for.⁶³ The association of modal and passive mood discreetly implies the reader’s engagement with the reading process. It is used only once more, to remind the reader that she knows more than the male character: “from a cause that may be readily conjectured, [Rachel] was more than usually silent in his company” (114). Such textual education partly relies on recognizing echoes with other texts, as will now be shown.

6.5 Intertextuality

The representation of reader agency comes as an imitation of Sterne’s openings. Taylor begins *Rachel* with the same variation on *Tristram Shandy*’s initial “I wish” that she had used two years earlier in *Display*:⁶⁴ “‘I wish I had not promised to go!’ exclaimed our hero, throwing aside a letter he had received by that morning’s post: ‘I shall be tired to death before the end of the week; but,’ continued he, after a long pause, ‘I will go, and who knows [...] but I may find a wife amongst them’” (*Rachel* 1). This is followed by an interruption from the reader which calls to mind the second sentence of *A Sentimental Journey* where an interlocutor intervenes,⁶⁵ while relying on one of the adjectives which will become a keyword for the heroine: “‘In the name of wonder, why then did he not marry?’ exclaims my fair reader. Alas! gentle lady, you have asked an awkward question. Be not then offended at the answer:—He was too old to lose his heart to a plume of white feathers, or a shortwaisted spencer. [...] But to return to our hero” (*Rachel* 1-2). “My fair reader” and the query by the “gentle lady” may also replay (in a toned down manner) the “silly question[s]” asked by Mrs Shandy (I.1) and the inquisitive female reader (I.4).⁶⁶ As anyone who has engaged with *Tristram Shandy* knows, female readers, especially if they are called “my fair Lady”—or a recognizable variation thereof—, should beware of being “inattentive in reading” (I.20). Taylor gives a somewhat more flattering version of a direct appeal to the reader: “Frown not, I beseech you, for in all probability he never knew *you*” (2). She illustrates her claim to have written “a plain story and probable events” (Preface) by having the reader interrupt the narrator a second time on the same page: “Tomkins, for that was our hero’s name — ‘*Tomkins!* what a hideous name!’” (2). This allows her to assert the narrator’s authority and sketch out the situation of a younger reader/listener: “—Hush! must I apologize twice in this

short space. I do confess it is neither romantic nor lover-like” (2).

The intertextuality of the heroine’s name may have escaped young readers but not their educators: “the name of Rachel [Clarence Hervey] could not endure” (*Belinda*, Ch. 26). By molding his protégée into the likes of “Virginia in M. de St Pierre’s celebrated Romance” *Paul et Virginie* (Ch. 26) and tapping into Rousseauist educational schemes, Edgeworth’s *Pygmalion* turns her away from him. Rather than being molded, Taylor’s Rachel grows into a character of which she gradually understands the benefits. Her evolution is in part presented with the help of a term which resonates within the entire corpus, and indeed, within the entire period covered by the three corpora and which will now be examined.

6.6 Sensibility

Rachel is the only text in the Women corpus after *Belinda* to use “sensibility” significantly more than a regular distribution of the term over the corpus would warrant, with 11 occurrences rather than the 2 expected.⁶⁷ Although only absent from three of the texts in this corpus,⁶⁸ 70% of its occurrences are located in or before Edgeworth’s 1802 novel, i.e., in the first half of the corpus ([figure 10a](#)). In the Reference corpus, the shift becomes clear after Burney’s 1814 *Wanderer* ([figure 10b](#)).⁶⁹

Taylor therefore resorts to a word which in 1817 had passed its heyday to model her heroine by contrast with other young females, finally making her male protagonist declare that “he was now convinced [Rachel] possessed as much sense, if not as much sensibility, as her cousin” (146), probably an allusion to Austen’s famous 1811 combination of these two words.⁷⁰ As Taylor always uses “sensibility” negatively, this amounts to a genuine compliment, particularly as Sophia is dismissed for “ill-directed sensibility” associated with selfishness and superficiality (100). As Todd points out, the “adjectives [associated with ‘sensibility’] tell the tale of its rise and fall” (7). What is developed here with *Rachel* is a case-study of a late use of the term.

The meaning of “sensibility” in Taylor’s text is explored through a discussion of the term by the characters. The uninteresting Miss Jenkins declares that what the other characters see as “an act of pure benevolence” is a proof that “sensibility may be reckoned almost a misfortune to its possessor,” which is judged as “sophistry” by Tomkins and reformulated as “true sensibility” (35). Immediately after this, words and acts are juxtaposed, triggering a change of perception:

“I hate sensibility in all its moods and tenses, as my cousin would say,” said Rachel, stooping to remove a snail which was crawling across the path.

“Surely the words are not suited to the deed,” thought Tomkins, as he regarded her with a degree of complacency he had never before felt towards her.

Miss Jenkins, who had not been very much pleased with his last observation, and willing to say any thing rather than remain silent, since it might lead him to suppose he had mortified her, exclaimed, “Bless me, Miss Porter, how can you endure to touch such vermin! I declare the sight of such reptiles always makes me shudder.”

“And does sensibility teach you to crush them to death with your foot, rather than save their lives by touching them with your finger?” asked Rachel. (35)

Tomkins's final positive judgment on *Rachel* capitalizes on the negative connotations which the word acquires along the text. Taylor's general strategy is to advocate for a middle ground of rational benevolent behavior, and she does this by debunking excessive and misapplied sensibility, which tallies with a low use of terms in the EMOTIONS list ([figure 5b](#)).

The early reception of the book shows that the characteristics of *Rachel* identified here were perceived by its contemporaries.

6.7 A "Plain Story"

The reviewer for *La Belle Assemblée* found the snail passage quoted above interesting enough to select it as one of the three short extracts illustrating the appreciation of *Rachel*, paying Taylor the (possible) compliment of taking her for a male author who, "like Sterne, seems fond of leaving something for the reader's imagination, for he does not inform us, though *Rachel* and Tomkins seem really to appreciate each other's good qualities, whether a second visit from the latter [...] will be productive of an indissoluble union." The *Monthly* reviewer's much shorter judgment points to the originality of Taylor's didacticism by disapproving it: "some of *Rachel*'s concealments may not be considered as offering a safe or useful example."⁷¹

Yet Taylor's intensive use of names⁷² draws out a safe and tightly-knit community in which the heroine can take risks. This is a successful example of small-group interaction identified in section 4 above as characteristic of female fiction, and qualified by Heuser and Le-Khac as "small, constrained social spaces" where "the legibility of people, their relationship, and their positions within the community" is of paramount importance, typical of early nineteenth-century fiction (33, 35). By incorporating the reader in the text as had been more customary in eighteenth-century fiction than when she wrote and by explicitly inciting the addressee to use her judgment, Taylor indeed provides a Johnsonian "lecture of conduct" (Johnson 176): the reader is in the comfortable position of knowing more than the characters and being capable of making the right inferences, an attitude which the *Monthly* reviewer did not condone.

Taylor's "plain story," as she puts in in her Preface (6), relies mostly on vocabulary familiar to her readers, used in unexceptionable frequencies, but sharply focused. The few words which are noticeably underused point to simple, unconvoluted syntax: fewer occurrences of "and" and "which" are to be found than the norm of the Women corpus ("negative specificities") – but here again, Taylor is moderate ([figure 11](#)).

Taylor's intelligent didacticism is encapsulated in the two lexical words *Rachel* stands out for: "apron" is distinctly overused (6 of the 32 occurrences in the Women corpus) and "love" underused (3 rather than the 24 expected). The garment emblemizes *Rachel*'s lack of preoccupation with conventions, which she discards at the end: the last use of the term reads "three several times she had been seen to take off her apron before she entered the parlour, when she knew Tomkins was in it" (140). While seemingly playing along with the rules of courtship narratives, and twisting a few of them (the "neither romantic nor lover-like name of the suitor," 2, and the lack of closure), the text centers less on the love plot than on the process of questioning, leading to maturity, which was shown to be a salient topic. In this lies the seeming

paradox of a highly original tale told in unoriginal terms, brought to light through a combination of close and distant reading, analyzing a series of “small touches” (Burrows, “Style” 186) in the perspective of several corpora.

The point of this case study is not to claim that *Rachel* should be acknowledged as a text having left a big footprint (Brewer). Rather, through the computer-aided analysis of vocabulary use, it illustrates some of the strategies by which the didactic potential of the courtship novel, and more widely of domestic drama, is made effective. *Rachel*’s few original traits make it a pleasant and therefore efficient variation on a well-established model. By grafting the debate about intelligence on the well-known plot and exploring ubiquitous terms such as *sensibility*, it provides a good illustration of how the suspense, or at least the interest, comes not from the ending of the story but from the process of discovery, “the inward progress towards judgement” which forms the “critical action” of “the ‘feminine’ novel” (Butler 145). Because Taylor addresses a young audience, she sticks to an unoriginal or standard use of vocabulary within the genre she adopted as a vehicle. *Rachel*’s typicality therefore contributes to making explicit for twenty-first century readers the norms underlying statements which were implicit for avid early nineteenth-century novel readers, such as Emma Woodhouse saying “just what she ought, of course. A lady always does” (Austen, *Emma* III.13).

7. Strategies of computer-aided textual analysis

Teasing out the complex web of typicality and atypicality which characterizes Taylor’s *Rachel* serves as an example to illustrate the sort of analyses that can be conducted with corpora built from and around texts like the Chawton novels in digital form. A large amount of text set up into corpora with a unifying criterion such as a period or a genre provides “empirical evidence to support stylistic judgments” and “understand the linguistic devices used to achieve [a particular] effect” (Biber 17, 19). Information provided by the Chawton corpus on its own and embedded within the Women one is complementary, and becomes more objective when compared to that obtained from the Reference corpus, which overlaps with the latter.

Zooming in on one novel or on a part of a corpus (texts written by female authors in the Reference corpus, for instance) makes it possible to use close reading techniques and identify very precisely the means used for one particular effect in an overall strategy: in the case study, Taylor’s didactic writing through the paradox of unoriginal vocabulary for a “highly original tale,” or, in the control corpus, the prevailing use of dialogue as a way for women writers to animate their networks of characters, akin to Austen’s “three or four families in a Country Village” (*Letters* 287).

Zooming out from a text into the corpus which includes it or from one corpus into a larger one provides a perspective from which to examine the use and effects of particular terms or set of terms. For instance, the corpora set up here provide an extensive ground on which to analyze the variations in meaning and in use of such a quantitative and qualitative “key term of the period” as *sensibility* (Todd), the “flexibility” of which has long been recognized (Barker-Benfield xviii). As Ruth Yeazell remarks about *modesty*, “all words so charged with value resist sharp definition, tend to surround themselves with fields of conflict and confusion” (6). Comparing and contrasting the ways in which such terms are used in a large and coherent body of text also makes it possible to

understand their many facets.

Other developments in the “Strategies of Writing” project will consist in analyzing aspects of the genre as a whole. Thematic research will consist in looking at the use of terms related to the body and to physiological reactions, the ways in which the “heroines blush, swoon” (Raven) and “cry”, or cry out, in relation to the EMOTIONS list and the influence of physiognomy. More linguistic research will delve into the use of parts of speech rather than single or thematically grouped words: for this more technical approach, *Hyperbase* will have to be relinquished in favor of, for example, *TXM*. The challenge, as all scholars exploring this new way of apprehending literary texts experience, is always to “mobilize quantitative data to make qualitative claims” (Heuser and Le-Khac 30).

In corpus linguistics, increasing the size of the corpora is always an advantage when typicality of usage is the main target. Distant reading of literature *à la* Moretti seeks a huge scale and can produce information such as the prevalence of a genre within the entire production of period (*Distant Reading*), genre classification from “unsupervised statistical analysis” (Allison et al.) and “quantitative literary history of 2,958 nineteenth-century British novels” (Heuser and Le-Khac). Garside, Raven and Schöwerling aimed at exhaustiveness for a given type of text within a distinct period from a bibliographical point of view. These lists, statistics and results have proved extremely valuable for the present study.

In corpus stylistics however, or in computer-aided textual analysis as I prefer to call it, the corpora must remain within a scope that a reader can encompass: large enough so that there is a sufficient “shared pool of stylistic and thematic materials” from which each author “pulls” (Jockers 162) but not so big that the internal logic of each text is irrelevant. A constant back and forth movement from the corpus to specific passages, from synthetic presentation of huge amounts of data to detailed lists of occurrences and the text itself, is what makes it possible to address the challenge of identifying features which constitute typicality (for instance of “the ‘feminine’ novel”) by the use of comparison, contrast, and refraction.

Acknowledgements

I would like to thank Chawton House Library for the wonderful work they do and for having granted me a visiting fellowship in April 2013. My gratitude also goes to the several readers who have helped improve earlier drafts of this essay.

ADDITIONAL FIGURES

Figure 3 – Words whose relative frequency is higher in texts written by female authors than by male authors (CNTROL34): top of the “positive specificity” list established by Hyperbase (left) exported into Excel (right)

(326 words in all, 266 without names)

écart	corpus	texte	mot	HIERAR	POSITIF
37.58	1529	1523	helen		
37.58	1497	1496	emmeline		
37.58	1252	1252	adeline		
37.58	1246	1246	delacour		
37.58	1219	1218	cecilia		
37.58	1153	1128	ellis		
37.58	1118	1117	juliet		
37.58	1016	1016	delamere		
37.58	977	976	emma		
37.58	971	968	fanny		
37.58	949	949	belinda		
37.58	934	934	montreville		
37.58	924	924	betsy		
37.57	56675	45923	her		
37.57	43315	36227	she		
37.57	9615	7876	lady		
37.57	8196	6819	miss		
37.57	6937	5803	mrs		
37.57	4146	3588	herself		
37.57	2217	2121	laura		
37.39	41516	26211	you		
36.03	3792	2999	cried		
35.57	773	773	harleigh		
35.41	110230	66123	-		
32.70	667	667	davenant		
32.61	664	664	hervey		

CNTROL34	Female authors	56%	
56 675	45 923	81%	her
43 315	36 227	84%	she
9 615	7 876	82%	lady
8 196	6 819	83%	miss
6 937	5 803	84%	mrs
4 146	3 588	87%	herself
41 516	26 211	63%	you
3 792	2 999	79%	cried

Figure 4 – Words whose relative frequency is lower in texts written by female authors than by male authors (CNTROL34): top of the “negative specificity” list established by Hyperbase

(67 words in all)

écart	corpus	texte	mot	HIERAR	NEGATIF
-37.55	217241	104872	the		
-37.55	89135	44774	a		
-36.49	46468	22547	his		
-35.67	141241	72959	of		
-35.39	5841	2111	our		
-33.69	138822	71933	and		
-29.40	10104	4347	we		
-28.53	1627	414	master		
-24.75	7134	3050	man		
-21.78	2080	716	men		
-19.45	3013	1195	old		
-18.98	1125	340	fellow		
-18.89	69171	36236	in		
-17.94	25873	13082	me		
-17.77	4157	1791	us		
-16.30	4539	2021	like		
-12.78	1751	723	three		
-12.16	1621	671	country		
-12.11	15300	7807	an		
-11.81	1872	800	gentleman		
-11.80	13971	7115	they		
-11.22	1339	550	captain		
-10.97	27750	14573	which		
-10.61	53644	28678	he		
-10.52	9424	4758	there		
-10.26	19927	10398	on		
-10.16	1065	434	small		
-10.11	991	399	case		
-9.96	1119	463	god		

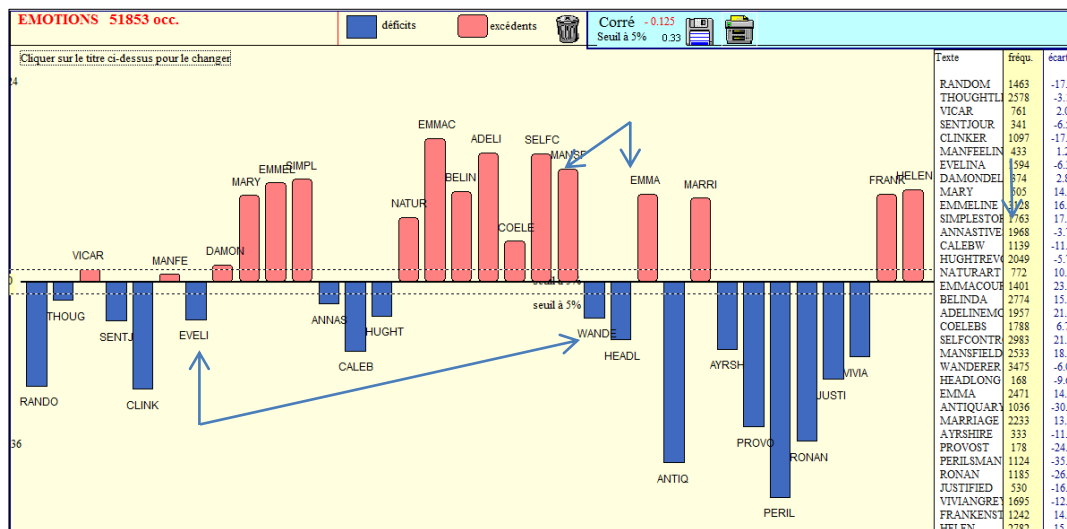
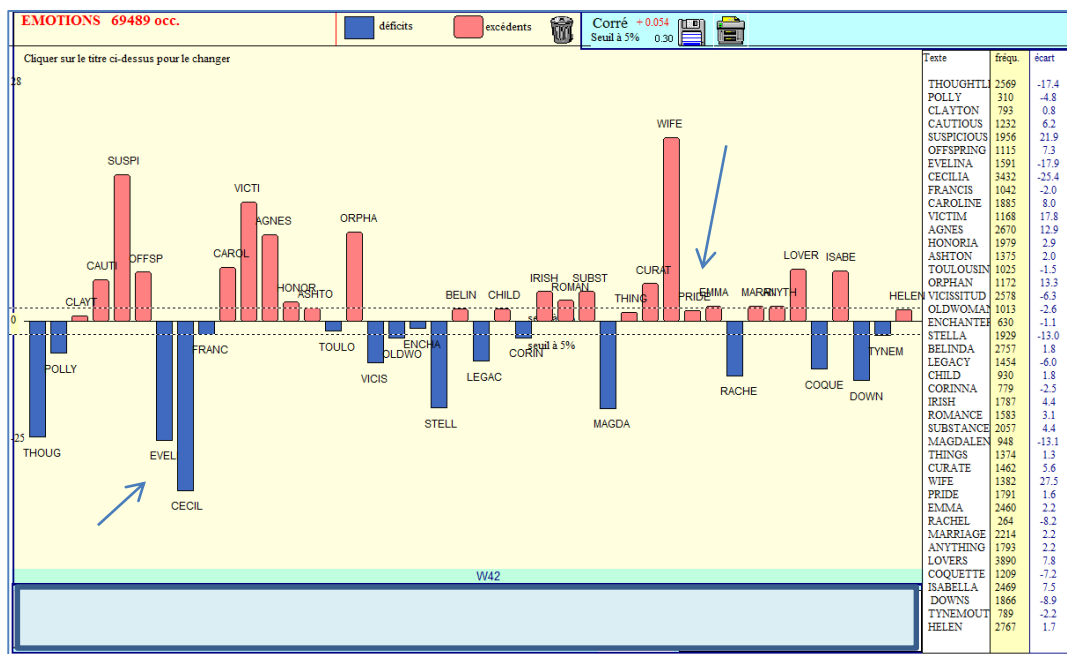
Figure 5a – EMOTIONS list in the Reference corpus (*Hyperbase*)Figure 5b – EMOTIONS list in the Women corpus (*Hyperbase*)

Figure 7– lexical connection on tokens in The Women corpus (*Hyperbase*)

Canonical novels in black, Minerva novels in blue, non-Minerva Chawton novels in red

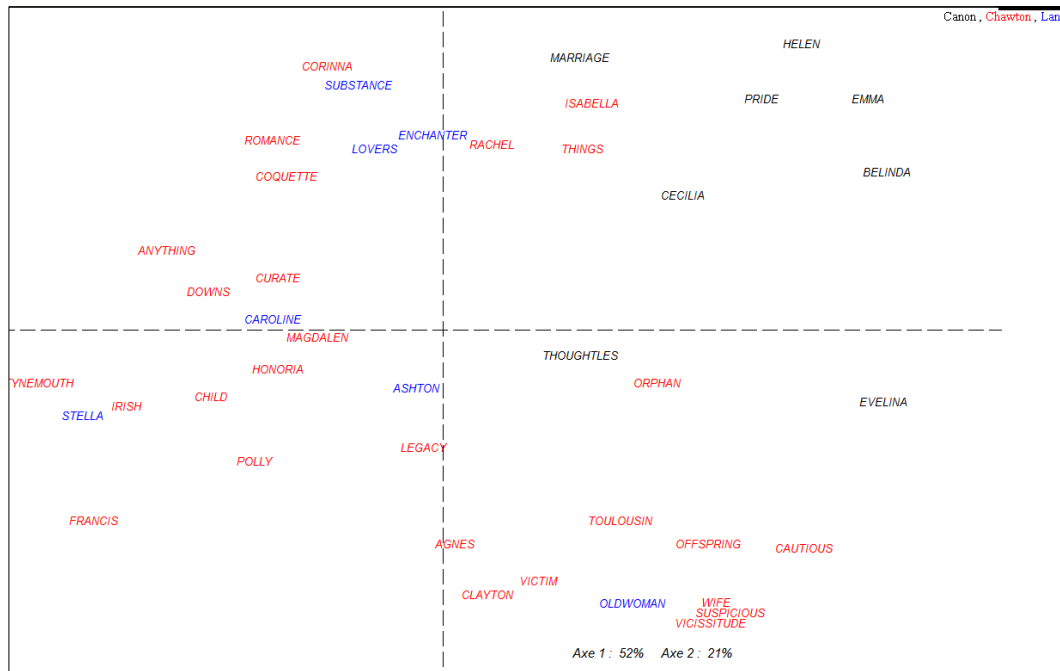


Figure 8 – lexical connection on tokens in CHAWTN34 (*Hyperbase*)

Minerva texts in black

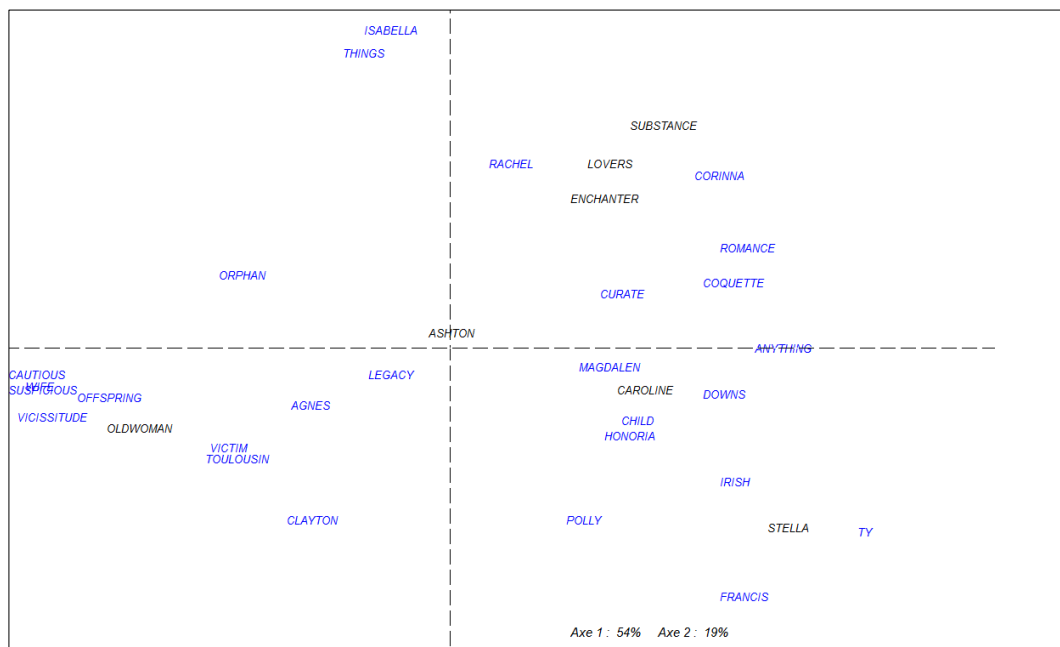


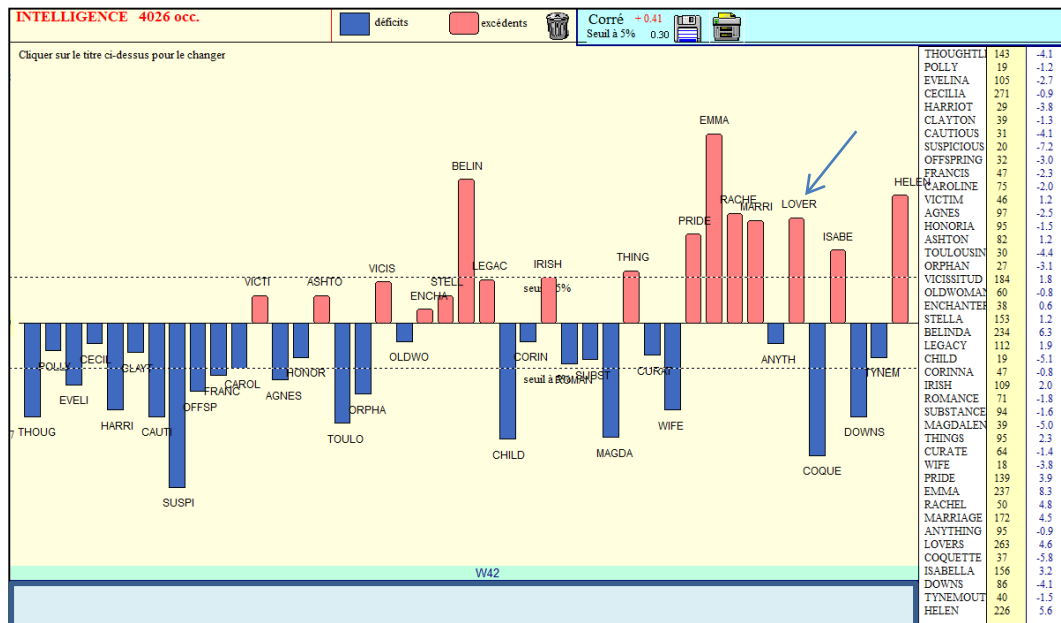
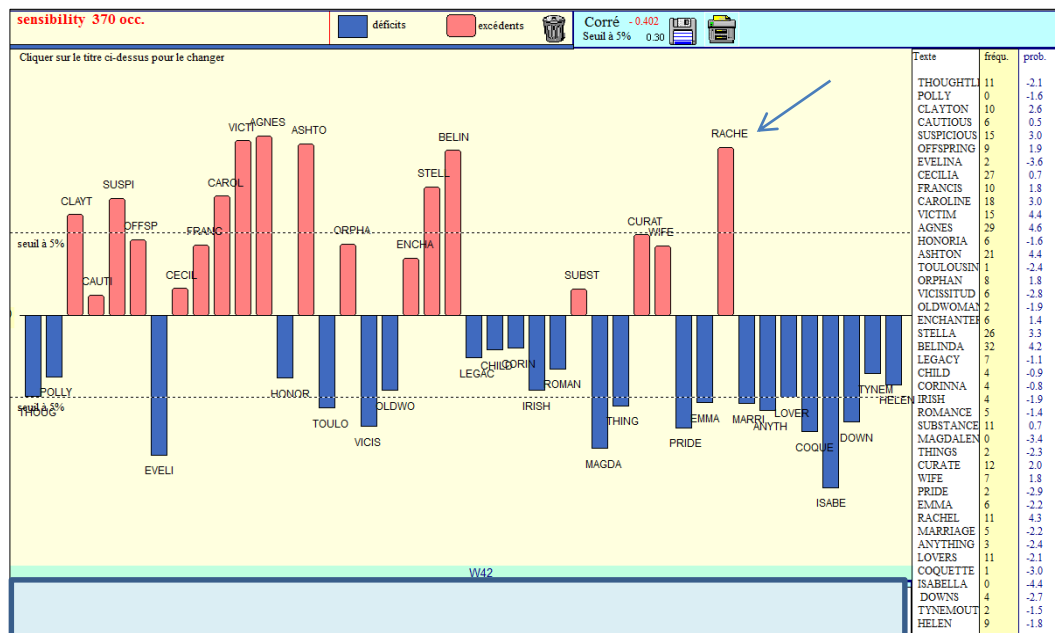
Figure 9 –Distribution of INTELLIGENCE as a lexical field in the Women corpus (*Hyperbase*)

Figure 10a –Distribution of “sensibility” in the Women corpus (*Hyperbase*)

1750 1778 1788 1794 1802 1814↓ 1815↑ 1823 1834

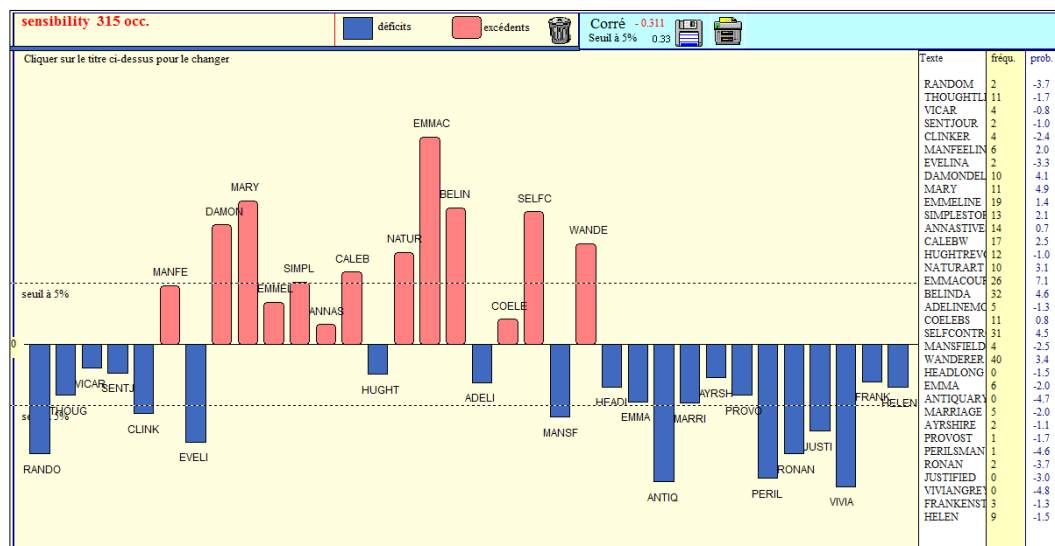
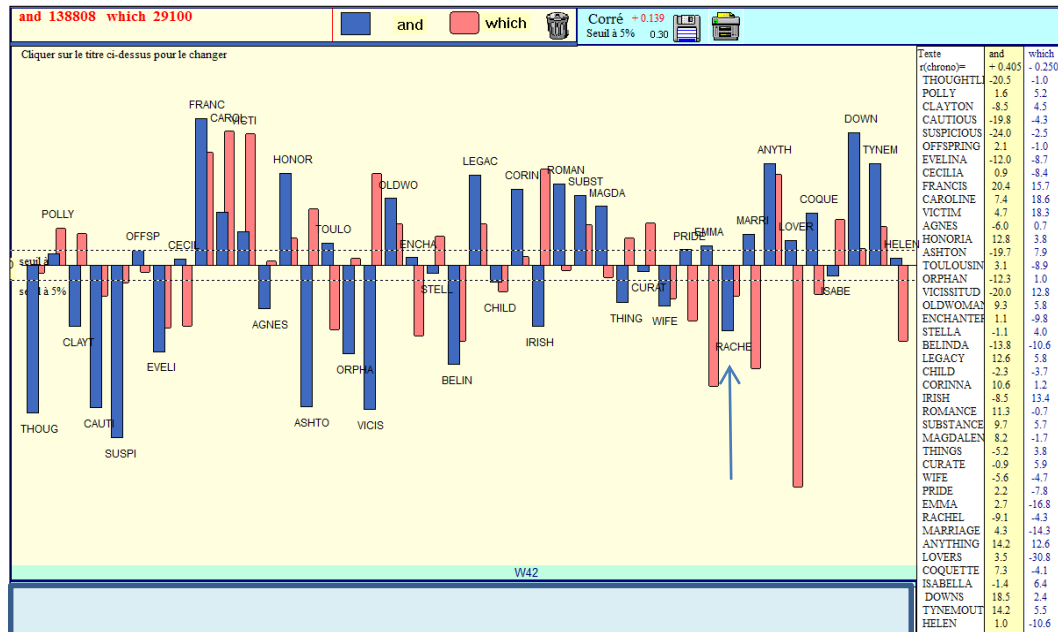
Figure 10b –Distribution of “sensibility” in the Reference corpus (*Hyperbase*)

Figure 11 – distribution of “and” and “which” in the Women corpus (Hyperbase)



THE CHAWTON CORPUS : CHAWTN34				
DATE	AUTHOR	TITLE	SHORT TITLE	corpu
1769	ANON	The Reward of Virtue; or, the History of Miss Polly Gra	POLLY	0,8%
1771	ANON	The History of Lord Clayton and Miss Meredith	CLAYTON	1,7%
1772	ANON	The Cautious Lover	CAUTIOUS	2,2%
1777	ANON	The Suspicious Lovers	SUSPICIOUS	2,6%
1778	CANNING	The Offspring of Fancy	OFFSPRING	1,9%
1786	JOHNSON	Francis, the Philanthropist: an unfashionable tale	FRANCIS	2,3%
1787	HUGUES	Caroline; or, the Diversities of Fortune	CAROLINE	3,3%
1787	TOMLINS	The Victim of Fancy	VICTIM	1,5%
1789	BENNETT	Agnes De-Courci: a Domestic Tale	AGNES	4,5%
1789	PURBECK	Honoria Sommerville	HONORIA	3,9%
1792	ANON	Ashton Priory	ASHTON	2,8%
1792	ANON	Prepossession; or, Memoirs of Count Touloussin	TOULOUSIN	2,2%
1793	MATTHEW	Simple Facts or the History of an Orphan	ORPHAN	1,7%
1794	LEWIS	Vicissitudes in Genteel Life	VICISSITUDES	6,1%
1800	CARVER	The Old Woman	OLDWOMAN	2,3%
1801	MARTIN	The Enchantress; or, Where Shall I Find Her? A Tale	ENCHANTERES	1,4%
1802	CRAIK	Stella of the North, or the Foundling of the Ship	STELLA	5,3%
1804	HUNTER	The Unexpected Legacy	LEGACY	3,6%
1808	WILKINSON	The Child of Mystery	CHILD	1,9%
1809	FOSTER	The Corinna of England, and a Heroine in the Shade: a N	CORINNA	1,8%
1809	MACKENZIE	The Irish Guardian, or, Errors of Eccentricity	IRISH	3,4%
1810	GREEN	Romance Readers and Romance Writers: a Satirical Na	ROMANCE	3,1%
1812	FOSTER	Substance and Shadow; or, the Fisherman's Daughter o	SUBSTANCE	4,0%
1812	HELME	Magdalen; or, the Penitent of Godstow	MAGDALEN	2,9%
1812	JACSON	Things By Their Right Names	THINGS	2,8%
1813	SPENCE	The Curate and His Daughter, a Cornish Tale	CURATE	2,7%
1813	COOPER	The Wife; or, Caroline Herbert	WIFE	1,4%
1817	TAYLOR	Rachel: a Tale	RACHEL	0,9%
1819	HARVEY	Any Thing But What You Expect	ANYTHING	3,6%
1821	HATTON	Lovers and Friends; or, Modern Attachments	LOVERS	7,3%
1822	ANON	The Village Coquette	COQUETTE	3,1%
1823	JACSON	Isabella. A Novel	ISABELLA	4,5%
1829	ANON	Three Weeks in the Downs, or Conjugal Fidelity Rewar	DOWNS	4,7%
1830	HARVEY	The Castle of Tynemouth. A Tale	TYNEMOUTH	1,8%
38% pre1800, 62% post1800			3 964 201	

THE WOMEN CORPUS: W42				
DATE	AUTHOR	TITLE	SHORT TITLE	% corpus
1752	Haywood	<i>Betsy Thoughtless</i>	THOUGHTLESS	4,9%
1769	ANON	<i>The Reward of Virtue; or, th</i>	POLLY	0,6%
1771	ANON	<i>The History of Lord Clayton</i>	CLAYTON	1,1%
1772	ANON	<i>The Cautious Lover</i>	CAUTIOUS	1,5%
1777	ANON	<i>The Suspicious Lovers</i>	SUSPICIOUS	1,8%
1778	CANNING	<i>The Offspring of Fancy</i>	OFFSPRING	1,3%
1778	BURNEY	<i>Evelina</i>	EVELINA	3,3%
1782	BURNEY	<i>Cecilia</i>	CECILIA	7,0%
1786	JOHNSON	<i>Francis, the Philanthropist:</i>	FRANCIS	1,6%
1787	HUGUES	<i>Caroline; or, the Diversities</i>	CAROLINE	2,3%
1787	TOMLINS	<i>The Victim of Fancy</i>	VICTIM	1,0%
1789	BENNETT	<i>Agnes De-Courci: a Domest</i>	AGNES	3,0%
1789	PURBECK	<i>Honorina Sommerville</i>	HONORIA	2,7%
1792	ANON	<i>Ashton Priory</i>	ASHTON	1,9%
1792	ANON	<i>Prepossession; or, Memoirs</i>	TOULOUSIN	1,5%
1793	MATTHEWS	<i>Simple Facts or the History</i>	ORPHAN	1,2%
1794	LEWIS	<i>Vicissitudes in Genteel Life</i>	VICISSITUDES	4,2%
1800	CARVER	<i>The Old Woman</i>	OLDWOMAN	1,6%
1801	MARTIN	<i>The Enchantress; or, Where</i>	ENCHANTERESS	0,9%
1802	CRAIK	<i>Stella of the North, or the F</i>	STELLA	3,6%
1802	Edgeworth	<i>Belinda</i>	BELINDA	3,9%
1804	HUNTER	<i>The Unexpected Legacy</i>	LEGACY	2,4%
1808	WILKINSON	<i>The Child of Mystery</i>	CHILD	1,3%
1809	FORSTER	<i>The Corinna of England, and</i>	CORINNA	1,2%
1809	MACKENZIE	<i>The Irish Guardian, or, Error</i>	IRISH	2,3%
1810	GREEN	<i>Romance Readers and Rom</i>	ROMANCE	2,1%
1812	FORSTER	<i>Substance and Shadow; or,</i>	SUBSTANCE	2,7%
1812	HELME	<i>Magdalen; or, the Penitent</i>	MAGDALEN	2,0%
1812	JACSON	<i>Things By Their Right Name</i>	THINGS	1,9%
1813	SPENCE	<i>The Curate and His Daughte</i>	CURATE	1,8%
1813	COOPER	<i>The Wife; or, Caroline Herb</i>	WIFE	1,0%
1813	Austen	<i>Pride & Prejudice</i>	PRIDE	2,5%
1816	Austen	<i>Emma</i>	EMMA	3,4%
1817	TAYLOR	<i>Rachel: a Tale</i>	RACHEL	0,6%
1818	Ferrier	<i>Marriage</i>	MARRIAGE	3,1%
1819	HARVEY	<i>Any Thing But What You Ex</i>	ANYTHING	2,5%
1821	HATTON	<i>Lovers and Friends; or, Mod</i>	LOVERS	5,0%
1822	ANON	<i>The Village Coquette</i>	COQUETTE	2,1%
1823	JACSON	<i>Isabella. A Novel</i>	ISABELLA	3,1%
1829	ANON	<i>Three Weeks in the Downs</i>	DOWNNS	3,2%
1830	HARVEY	<i>The Castle of Tynemouth. A</i>	TYNEMOUTH	1,2%
1834	Edgeworth	<i>Helen</i>	HELEN	3,9%
41% pre 1800, 59% post1800			5 823 537	
Blue: also in CHAWTN34				

THE REFERENCE CORPUS CTROL34					
DATE	AUTHOR		TITLE	SHORT TITLE	% corpus
1748	Smollett, Tobias	m	<i>Roderick Random</i>	RANDOM	4,1%
1752	Haywood, Eliza	fe	<i>Betsy Thoughtless</i>	THOUGHTLES	5,3%
1766	Goldsmith, Oliver	m	<i>Vicar of Wakefield</i>	VICAR	1,4%
1768	Sterne, Laurence	m	<i>Sentimental Journey</i>	SENTJOUR	0,9%
1771	Smollett, Tobias	m	<i>Humphrey Clinker</i>	CLINKER	3,3%
1771	Mackenzie, Henry	m	<i>Man of Feeling</i>	MANFEELING	0,8%
1778	Burney, Frances	fe	<i>Evelina</i>	EVELINA	3,6%
1784	Godwin, William	m	<i>Damon & Delia</i>	DAMONDELIA	0,6%
1788	Wollstonecraft, Mar	fe	<i>Mary</i>	MARY	0,5%
1788	Smith, Charlotte Tur	fe	<i>Emmeline</i>	EMMELINE	4,6%
1791	Inchbald, Elizabeth	fe	<i>A Simple Story</i>	SIMPLESTORY	2,3%
1792	Holcroft, Thomas	m	<i>Anna St Ives</i>	ANNASTIVES	4,1%
1794	Godwin, William	m	<i>Caleb Williams</i>	CALEBW	3,0%
1794	Holcroft, Thomas	m	<i>Hugh Trevor</i>	HUGHTREVOR	4,4%
1796	Inchbald, Elizabeth	fe	<i>Nature and Art</i>	NATURART	1,0%
1796	Hays, Mary	fe	<i>Emma Courtney</i>	EMCOURTNEY	1,5%
1802	Edgeworth, Maria	fe	<i>Belinda</i>	BELINDA	4,1%
1804	Opie, Amelia	fe	<i>Adeline Mowbray</i>	ADELMOWBRAY	2,4%
1809	More, Hannah	fe	<i>Coelebs</i>	COELEBS	3,0%
1810	Brunton, Mary	fe	<i>Self-Control</i>	SELFCONTROL	4,0%
1814	Austen, Jane	fe	<i>Mansfield Park</i>	MANSFIELDP	3,5%
1814	Burney, Frances	fe	<i>Wanderer</i>	WANDERER	7,4%
1815	Peacock, Thomas Lo	m	<i>Headlong Hall</i>	HEADLONG	0,6%
1816	Austen, Jane	fe	<i>Emma</i>	EMMA	3,6%
1816	Scott, Walter	m	<i>Antiquary</i>	ANTIQUARY	4,0%
1818	Ferrier, Susan	fe	<i>Marriage</i>	MARRIAGE	3,3%
1821	Galt, John	m	<i>The Ayrshire Legatees</i>	AYRSHIRE	1,1%
1822	Galt, John	m	<i>The Provost</i>	PROVOST	1,2%
1822	Hogg, James	m	<i>The Three Perils of Ma</i>	PERILSofMAN	4,7%
1823	Scott, Walter	m	<i>St Ronan's Well</i>	RONAN	4,1%
1824	Hogg, James	m	<i>Justified Sinner</i>	JUSTIFIEDSIN	1,8%
1826	Disraeli, Benjamin	m	<i>Vivian Grey</i>	VIVIANGREY	4,3%
1831	Shelley, Mary	fe	<i>Frankenstein</i>	FRANKENSTEIN	1,6%
1834	Edgeworth, Maria	fe	<i>Helen</i>	HELEN	4,1%
				5 481 755	
Green: also in W42		56% female	41% pre1800		
		44% male	59% post1800		

Notes

¹ Aphra Behn's 1729 *The Rover; or, the Banish'd Cavaliers* and Penelope Aubin's 1723 *The Life of Charlotta Du Pont*, brazenly reissued in 1770 as *The Inhuman Stepmother, or the History of Miss Harriot Montague*, also in the collection (see Kulik).

² The chapter has not been kept in the 2010 *Cambridge Companion to Jane Austen*. I have not had access to *Graphing Austen: The Evolutionary Basis of Literary Meaning* by Joseph Carroll, Jonathan Gottschall, John A. Johnson, and Daniel J. Kruger (Palgrave Macmillan, 2012), but the description of the book indicates a rather different focus—and a somewhat misleading title: “Hundreds of readers gave numerical ratings to the attributes of hundreds of characters (Austen to Forster) and also rated their own emotional responses to the characters. We draw conclusions about the determinacy of literary meaning, interactions between gender and the ethos of community, and the adaptive function of morally polarized characterization” (<http://www.ebooks.com/956586/graphing-jane-austen/carroll-joseph-gottschall-jonathan-johnson-john-a/>).

³ “A concordance can present and sort data in a form which allows human analysts to see patterns more easily” (Ho 201).

⁴ Most studies in English rely at least partly on *Wordsmith Tools*. My main tool is *Hyperbase: Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels*, version Windows 9.0. INALF, 2010. Although devised for French, it can be used with any language that uses a roman alphabet and it provides reference corpora for English, German, Spanish, and Italian. As the reference corpus for English is the [British National Corpus](#) (“a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century”), it was deemed necessary for this study to set up a corpus of texts contemporary to the Chawton novels to use as reference (see section 2.3).

⁵ J.C.T. Oates called the Sterneana he collected “the rubbish of literature” but treasured it for the access it gives us to “the vogue for things Shandean and sentimental” (4). My first experiments with computer-aided textual analysis compared the use of the most frequent words in the nine authentic and the two spurious volumes in *Tristram Shandy* (“*Tristram Shandy*: Créations et imitations en Angleterre au XVIIIe siècle”, doctoral dissertation, Paris Sorbonne Nouvelle,

1991): from the list of word frequencies I showed some of the reasons why Spurious Volume 3 was immediately spotted as such while Spurious Volume 9 fooled quite a few readers, including a translator. A more developed version appears in Bandry-Scubbi ("Les livres").

⁶ François Rastier opens his 2011 *La Mesure et le grain: Sémantique de corpus* with the anecdote of answering a journalist's question "what is love" by "in which corpus?"; the question was probably triggered by the title of his 1995 *L'analyse des données textuelles: l'exemple des sentiments dans le roman français 1820-1970* (11).

⁷ For work on smaller corpora, see Bandry-Scubbi "Les livres," "Gulliver," "Roderick Random," "Renaissance de/chez," "Evelina," "Yes, Novels," "Space and Emotions," and "Chawton House Library").

⁸ This criteria eliminates thirteen texts: [*A New Atalantis*](#), [*The Death of Grimaldi, or the Fatal Secret, De Montmorency: a Novel Founded on Recent Fact*](#), [*The Imposters Detected or the Life of a Portuguese*](#), [*The Monk and the Vine-dresser: or, the Emigrants of Bellesme*](#), [*Yamboo; or, the North American Slave*](#), [*The Parisian; or, Genuine Anecdotes of Distinguished and Noble Characters*](#), [*A Peep at the Pilgrims*](#), [*The Princess of Cleves*](#), [*Villasantelle; or the Curious Impediment*](#), [*Cava of Toledo; or, the Gothic Princess*](#), [*Drelincourt and Rodalvi; or, Memoirs of Two Nobel Families*](#), and [*Paris Lions and London Tigers*](#).

⁹ The two historical texts are [*The Lord of Hardivyle, an Historical Legend of the Fourteenth Century*](#) and [*Monmouth: a Tale, Founded on Historic Facts*](#).

¹⁰ This criteria eliminates [*The Cruel Husband; or, Devonshire Tragedy*](#) (1,900 words), [*The History of Miss Sally Johnson; or, the Unfortunate Magdalen*](#) (8,000 words), and [*Drawing-Room Tales. The Stout Gentleman; The Deserter; and The Broken Heart*](#) (a series of anecdotes). The two early texts and the reprint mentioned in note 1 above were not included either.

¹¹ The term "word" calls for a definition, which varies according to the software used and causes slight differences in global counting. The definition adopted for this study (and in all my computer-aided work) is basic and straightforward: a string of characters surrounded by blanks or punctuation marks. Lemmatization, or the grouping together of words that are inflected or variant forms of the same word, is often, but not always, useful.

¹² The number of texts is included in the names of the corpora so as to leave the possibility of working with other versions at a later stage of the project and making the differences clear. So as not to swamp the reader of the present essay with figures, the more general denomination of the "Chawton corpus," the "Women corpus" and the "Reference corpus" are used.

¹³ The Chawton Minerva texts are Anne Hughes's [*Caroline; or, the Diversities of Fortune*](#); [*Ashton Priory*](#); Mrs. Carver's [*The Old Woman*](#); Mrs Martin's [*The Enchantress; or, Where Shall I Find her? A Tale*](#); Helen Craik's [*Stella of the North, or the Foundling of the Ship*](#); Mrs. E.M. Foster's [*Substance and Shadow; or, the Fisherman's Daughter of Brighton*](#); Anna Julia Kemble Hatton's [*Lovers and Friends; or Modern Attachments*](#).

¹⁴ The Chawton epistolary novels are [*The Suspicious Lovers*](#); Mary Ann Canning's [*The Offspring of Fancy*](#); Elizabeth Sophia Tomlins' [*The Victim of Fancy*](#); Anna Maria Bennett's [*Agnes De-Courci*](#); [*The Cautious Lover*](#); Alethea Lewis' [*Vicissitudes in Genteel Life*](#); Mrs. Carver's [*The Old Woman*](#); Maria Susanna Cooper's [*The Wife; or, Caroline Herbert*](#).

¹⁵ As Fischer-Starcke notes, "legal access to electronically stored language data is one of the necessary preconditions for corpus linguistic and corpus stylistic analyses" (8). In the ten years or more since she set up her corpora for *Jane Austen and her Contemporaries*, published in 2004, Project Gutenberg has developed tremendously both in scope and in quality.

¹⁶ This 1860-1960 corpus was established by my doctoral student Caroline Orbann as a reference corpus for a study of children's literature 1863-1973. It comprises 15 texts by Dickens, Eliot, Hardy, McDonald, Conrad, Forster, Lawrence, Woolf, Orwell, Bowen, Lessing, and Dahl, which add up to 1.3 million words.

¹⁷ The software examines the presence and absence of every word in any combination of two texts in the corpus and processes this huge amount of data (a correlation matrix) with the data reduction technique of principal component analysis "to see whether there is a set of factors that can explain the variation of the variables under study" (Hinton 305). "A principal component ranks the likelihood of certain features occurring, so texts are sorted according to the features they lack, as well as by the features they share" (Allison et al. 20). *Hyperbase* proposes maps of factors 1 and 2 ("Axes 1 & 2" in the figures), 1 and 3, 2 and 3; because the first two factors carry the most weight, their representation is the most useful. A very simple example of such a matrix is the distance chart usually found in road atlases (with only two dimensions). A more elaborate yet easy to understand explanation can be found [here](#) (Dallas).

¹⁸ Word-types are the different words or character-strings of the text (which form the vocabulary) for which only presence or absence counts. Words are considered as tokens when their frequency is taken into account. See for example the explanation in Wordsmith Tools: http://www.lexically.net/downloads/version5/HTML/index.html?type_token_ratio_proc.htm.

¹⁹ The horizontal axis in figure 1 calculates the first factor which carries 52% of the "weight" of data. This means that respective positions along this line are far more meaningful than those along the vertical axis (second factor, 14%): for example, the proximity of *Emma* and *Mansfield Park* to *Evelina* and *Emma Courtney* is much stronger than their difference or than the proximity of *Belinda* and *Coelebs* to *The Provost* and *The Ayrshire Legatee*. Holcroft taps into the common vocabulary of female texts, Godwin does so more for *Caleb Williams* than for *Damon and Delia*.

²⁰ *Betsy Thoughtless* is close to them according to the first factor only (i.e., closer than texts focusing on male characters, which is not a surprise), and this is also the case in figure 7. The explanation may reside in chronology but will not be envisaged in the present study. For its proximity to Haywood's earlier fiction see Bandry-Scubbi ("Renaissance de/chez").

²¹ Fischer-Starcke's "contemporary literature" corpus is made up of 30 canonical texts from *Pamela* to *Adam Bede*, adding up to "about 4 370 000" tokens (190-91). Graves somewhat optimistically aimed at establishing the "'vocabulary fingerprint' for a writer" ("Computer Analysis" 204).

²² I find "specific" and "specificity" in the sense of *Hyperbase*, and of French terminology more generally, intuitive (Lebart and Salem); it is the equivalent of quantitative "keyness" in studies which rely on *Wordsmith Tools* (such as most works dealing with corpus stylistics in English, notably those mentioned at the onset of this study). The *Hyperbase* calculation takes into account whether the word is very frequent or not, as this impacts expectations about the distribution. Keyness or specificity should not be confused with frequency.

²³ Fischer-Starcke documents this by commenting on research by Culpeper and Toolan (69). In most cases, the items at the top of "specificity lists" do not come as a surprise; rather than proving the method useless, this gives validity to further findings, as Mahlberg (2007) points out, quoting Stubbs: "even if quantification only confirms what we already know, this is not a bad thing. Indeed, in developing a new method, it is perhaps better not to find anything too new, but to confirm findings from many years of traditional study, since this gives confidence that the method can be relied on" (qtd. in Mahlberg, "Corpus Stylistics" 228).

²⁴ There are 104,393 occurrences of *she*, *her*, *herself*, *hers* in the Reference corpus, and 85,934 in the texts written by women taken as a whole. Burrows calls these terms "pronouns" despite their mixed grammatical status (*Computation into Criticism*). The proportion of male and female pronouns proved to be a discriminating factor between spurious continuations of *A Sentimental Journey*, as shown in Bandry-Scubbi ("Les livres"). For the stylistic effect of different pronouns in *Gulliver's Travels*, see Bandry-Scubbi ("Gulliver").

²⁵ Pronouns feature among "female indicators" in the study of "a genre-controlled subset of the BNC," a large corpus of fiction from 1960-1993 (Koppel, Argamon, and Shimoni section 4). Other gender features tally with the present findings, notably that determiners are among the "male indicators" (see figure 4).

²⁶ Other "scarce" words in the female part of the Reference corpus draw up a clichéd perception of male preoccupations in male texts: *country* and *God* (59% male), *ground* (57% male), *against* (54% male), *stood*. 57% for *stand* lemmatized, but 50% without *Perils of Man* where this word is overused with no specific pattern, as a concordance shows. This is the case even if the other nouns in the list of female "negative specificities" are too influenced by their strong presence in a single text to serve for generalizations: *order*, *spirit*, *act*, *justice*, *purpose*, *place*, *length*, *public*, *light*, *company*.

²⁷ Fischer-Starcke identifies and analyses "family or social relationships" as one of these patterns or "dominant topics" in Austen (95-105). Graves looks at "feelings" in *Emma* ("Computer Analysis" 209).

²⁸ “EMOTIONS” is taken here as an umbrella term. The 35 EMOTIONS words are: *admiration, affection, affections, angry, anxious, attachment, attention, comfort, dear, delight, esteem, feel, feeling, feelings, friendship, glad, gratitude, happy, heart, joy, kindness, loved, lover, pain, passion, pity, pleasure, sensible, smile, sorrow, sweet, tears, tender, tenderness, and unhappy*.

²⁹ The Women corpus has 11.76‰ while the Chawton one is at 12.59‰. In the 1860-1960 corpus the rate is 5.3‰ in male texts and 6.7‰ in female ones, which confirms the high presence of these terms in the earlier period which forms the object of this study.

³⁰ Time, also a pattern identified by the positive specificity list of female texts in the Reference corpus, would necessitate a more detailed study, as was conducted for *Roderick Random* (Bandy-Scubbi, “*Roderick Random*” 211-13).

³¹ For example, these terms include *ship, deck, water, board, sea, cape, officers, lieutenant, cabin, gunner, and sail*, etc.

³² Authorship studies form a major trend in quantitative textual studies. A few controversies have made this approach infamous (the Shakespeare authorship question, the Corneille / Molière argument after Dominique Labbé’s *Corneille dans l’ombre de Molière*), but it is also used in forensic linguistics. A less spectacular study on author-consistency focuses on the Romain Gary / Émile Ajar controversy (they were both awarded the Prix Goncourt before it was realized they were the same person whereas the prize can only be given to an author once): Vina Tirvengadam argues that authors can “consciously manipulate” their writing habits. There is no reason to think that authors examined in the context of this study willingly changed their writing habits, so it will be assumed, at least at this stage of the project, that thematic purposes justify differences between texts by the same novelist.

³³ “Wellhopper,” who has stopped at the end of the first volume, may refer to the Dodo Press *Ashton Priory*, or have downloaded the text on a portable reading device (<http://gothicheroine.wordpress.com/category/authors/anon/>). Contrary to *Powis Castle* (1788) and *Benedicta* (1791), the anonymous author’s previous works featured on the title-page, also from Minerva, and to the 1794 *Mariamme*, *Ashton Priory* is not in *ECCO* and it is listed in neither ESTC, NUC, nor Blakey. It has an entry in McLeod, 217. The versions which come up online duly refer to Chawton House Library.

³⁴ This is found in Austen’s famous defense of the novel, *Northanger Abbey* I.5.

³⁵ *Emma* is actually longer than *Legacy*; *Pride & Prejudice* about the same size. Austen’s texts have three volumes, not four—but *Legacy* has two.

³⁶ Because the factors’ difference in weight is lower for tokens than for types, the grouping of Austen, Burney, Edgeworth, and Ferrier in the same area of the graph is highly significant.

³⁷ Sarah Green, [*Romance Readers and Romance Writers: a Satirical Novel*](#) Number 9 of Chawton House Library Series, Pickering and Chatto, 2010.

³⁸ The entry also states, “Though the book reached a third edition, very few copies are now known” (Jane Taylor entry: Life & Writing screen within Brown, Clements, and Grundy. *The English Novel 1770-1829* records the three editions (the last in 1821) and another in Philadelphia (1818). *ODNB* does not mention *Rachel* but defines Jane Taylor as a “children’s writer” whose most famous text is “Twinkle, twinkle little star.” Isobel Grundy only mentions her as a poet (“(Re)discovering Women’s Texts” 182), as does the *OED* which only draws quotations from her 1816 *Essays in Rhyme* and from *Poems for Infant Minds* published in 1805 with her sister Ann Gilbert. *The History of British Women’s Writing 1750-1830* does not refer to Jane Taylor’s fiction (Labbe). Neither sister appears in the Chawton House Library [Biographies of Women Writers](#). A good electronic facsimile of the 2d 1818 edition of *Rachel* with a fancy but unreliable search function and a very poorly scanned plain text version are available from [https://openlibrary.org/ works/OL16288278W/Rachel](https://openlibrary.org/works/OL16288278W/Rachel). Several printed versions are advertised (most probably printed on demand, and from experience poorly so: DodoPress, Hardprinting, as of 10 June 2014). This makes the good quality of Chawton Novels Online all the more valuable.

³⁹ It is the most central text for its vocabulary (types, figure 6) and very close to the central axis for the frequencies in which it uses words (tokens, [figure 7](#)).

⁴⁰ *Rachel* is made up of 34,426 tokens and 3,794 types (as counted by *Hyperbase*). *Polly* has 33,584 tokens and 3,945 types. Their Type Token Ratios are very close (0.110 for *Rachel* and 0.117 for *Polly*). Incidentally, as comparing this ratio only makes sense for texts close in length, calling it “lexical richness” as Jockers does is somewhat misleading (54). As Cossette points out, the many ways proposed to measure lexical richness point to a complex problem (4).

⁴¹ *Cælebs in Search of a Wife* has been called “One of the most popular novels of the early nineteenth century” (Corman 24), and “spectacularly successful” (Waldron).

⁴² *Rachel* ranks as the fourth highest text in the corpus for the use of “appear” (lemmatized) proportionally to its size (0.6% of the Women corpus), with 64 occurrences rather than the 30 expected. All but 9 are in the sense of “seem” or, for the noun, “the action or state of appearing or seeming to be (to eyes or mind); semblance; looking like” (“Appearance,” def. 8). There is no difference in the quantitative use of “appear-” by male and female authors in the Reference corpus. 9 instances of “as if” does not differ significantly from the expected value of 7 for a text of this length in the Women corpus.

⁴³ This is the case when “foil” is lemmatized, as noun and verb. Several texts use “foil-” two or three times but variations in the distribution are not statistically significant. The slightly larger Reference corpus has 21 occurrences only, of which 13 are in novels by women (marginally higher than the proportion of text written by women); more appositely, these include the 3 instances which compare female characters.

⁴⁴ The anonymous [The Cautious Lover, or the History of Lord Woburn. By a Young Gentleman of Oxford](#) has 2 volumes.

⁴⁵ Her “drooping spirits” are revived by the compliments of a young man. Maria Susannah Cooper’s *The Wife; or Caroline Herbert* also gives a male version of the term in an interesting metaphor: “the fellow [described as “the most squeamish, sentimental mortal I ever met with”] draws his pen upon me, foils me desperately at this weapon” (Letter XXXV).

⁴⁶ Fanny refers to, “My foolishness and awkwardness,” which is taken up and qualified by Edmund (*Mansfield Park*, I.3). This comes after “her air, though awkward, was not vulgar” (I.2). Edward Ferrars is another self-reflexive user (“I am only kept back by my natural awkwardness”), while Colonel Brandon presents himself as “a very awkward narrator” when telling the story of Eliza to Elinor (*Sense and Sensibility* II.9). Catherine Morland’s “thin awkward figure” could not have been known to Taylor for chronological reasons (*Northanger Abbey* I.1), but “Mr. Collins, awkward and solemn” could have been known (I.18). More also makes the collocation self-reflexive, with Cœlebs as a first-person narrator being “chid for [...] my awkwardness” (Ch. 40).

⁴⁷ Waldron notes Austen’s “sardonic” comment on Hannah More’s text: “Of course I shall be delighted when I read it, like other people but till I do, I dislike it” (*Jane Austen’s Letters* 177). Olivia Murphy puts Austen’s reaction against *Cœlebs* in the wider perspective of her “point[ing] out the dangerous nonsense of conservative fiction” (298). With 5 occurrences of “awkward-” More conforms to the norm of the Reference corpus in terms of frequency.

⁴⁸ The only other text which uses “awkward” or “aukward” and their derivatives proportionately more frequently than the rest of the Women corpus is Anna Maria Mackenzie’s *The Irish Guardian, or Errors of Eccentricity* but it applies its 25 occurrences of the adjective, 4 of the adverb and 2 of the noun indifferently to characters, feelings or actions, as a concordance shows. 6 occurrences are expected for this text in the Women corpus, so Mackenzie’s use of “awkward” may reflect her subtitle, *Errors of Eccentricity*.

⁴⁹ He says, “Considering that it would be rather awkward to be clapped up just now” (65), and “It is rather awkward, to be sure. Why, I say, Rachel, what a crusty old woman that aunt of our’s must be! well! [...] what must be must, you know; so I shall take wing again; for stay here I cannot” (76).

⁵⁰ “Tall” is used in 91 occurrences for men, 52 for women, 12 for plants, and 8 for diverse items. The 2 occurrences in *Rachel* correspond to the expected quantity in the Women corpus.

⁵¹ The first example of “thick in the body, not lean or slender; usually in unfavourable sense, inclined to corpulence” (“Stout, def. 12a) in *OED* dates from 1804. A quotation from *Sense & Sensibility* illustrates the meaning of “strong in body; of powerful build”: “They must get a stout girl of all works” (qtd. in “Stout,” def. 6a). Austen clearly likes the word and uses it significantly more than the norm of the Women corpus would cause one to expect, at least once in each of her six novels and particularly in *Northanger Abbey* with 5 occurrences when at most 1 is expected for a text of this length (3 in both meanings for different characters, 2 with “heart”). In the Women corpus, “stout” comes with “heart” 6 times and with “fellow/s” 7 times out of the 36

uses. Only 2 of the 76 occurrences of “stout” in the Reference corpus apply to women; Walter Scott is the highest user (20 in *Vivian Grey* and 9 in *St Ronan’s Well*).

⁵² The *OED* defines “stupid” as: “Wanting in or slow of mental perception; lacking ordinary activity of mind; slow-witted, dull” (def. 3a).

⁵³ The vicar says, “She is one of those many insignificant beings, who bring so much discredit on the sex: her best qualification is, that she is too harmless to do much evil, and her intellect is too weak to suffer her to do much good” (Taylor 36).

⁵⁴ *OED* gives no example for “idiot” as “a person without learning; an ignorant, uneducated person; a simple or ordinary person” without historical reference after 1722 (def. 1a). The meaning of “a person so profoundly disabled in mental function or intellect as to be incapable of ordinary acts of reasoning or rational conduct” (def. 2a) is illustrated in its use as an adjective, “that is an idiot; profoundly mentally and intellectually disabled; exceedingly stupid,” by, among others, Wordsworth’s 1798 “The Idiot Boy” (def. B1).

⁵⁵ *Cælebs* is one of the three texts that stand out for their use of “Latin” in the Reference corpus, the other two being *Roderick Random* and *The Antiquary*. More jokingly links the knowledge of Latin by women to their inability to provide good meals (Ch. III, with the additional twist that the girls accused of knowing Latin in fact read sentimental novels, reversed in Ch. V: “There was not the least suspicion of Latin here”) before making it a part of her paragon Lucilla Stanley’s successful education: “A learned language, which a discreet woman will never produce in company, is less likely to make her vain than those acquirements which are always in exhibition” (Chapter XXXVIII); it is linked to male “prejudice” against female “scholars” and educated girls, which the author denounces with a note on Rev. Samuel Seyer’s 1808 *Latin Redivivum, or the modern use of the Latin Language, and the prevalence of the French* (footnote 4). The use of “Latin” also makes only three texts stand out in the Chawton corpus: of the 24 occurrences 4 are in *Rachel*, 7 in Elizabeth and Jane Purbeck’s *Honorina Sommerville* and 9 in Anna Julia Kemble Hatton’s *Lovers and Friends; or Modern Attachments*, while the other novels have 1 to none. All come with negative connotations.

⁵⁶ “Latin” and “idiot” are linked only once in *Lovers* however, when the young man who “had obtained several prize medals for his Greek and Latin orations” is later deemed “an absolute idiot in the customs of the world: in his person puritanical and formal, his language composed of far fetched words and Latin quotations.”

⁵⁷ There are 217 occurrences: 162 of the adjective, 52 of the noun, and 2 of the adverb. There are 3 occurrences in *Rachel*: more than the 1 to 2 expected but not to the extent of being statistically significant. The Reference corpus has proportionately slightly fewer occurrences of “stupid-” (187) and almost half of the 87 instances of “idiot” and “idiocy” are in the texts by Holcroft and Disraeli (dispersed as insults or self-imprecations in the texts of the first, mainly for high-ranking and mentally-disabled characters in *Vivian Grey*): this confirms the pointed use Taylor makes of the rarer term.

⁵⁸ There are 150 occurrences in WOMEN: 52 *intellect*, 47 *intellects*, 51 *intellectual*. *Rachel* ranks third for the relative frequency of this lemma, after *Isabella* and Ferrier's *Marriage* (this text being the same size as *Isabella* but having only 14 occurrences confirms the latter's intensive use of the lemma). Again, this points to the influence of More's *Coelebs*, which stands out strikingly in the Reference corpus for its frequent use of the lemma (43 rather than the 7 occurrences expected).

⁵⁹ Frances Jacson, *Isabella. A Novel*. This is the only association of "bewildered" with "intellect" in WOMEN: either as a verb or an adjective, "bewildered" collocates with words used to describe people in 2/3 of its 59 occurrences, and with "senses" (6), "brain" (4), "faculties" (4), "imagination" (2), ideas, thoughts, mind, head, heart (1 of each). Work on Frances Jacson is in progress.

⁶⁰ *Intellig-*, *intellect-*, *clever-*, *comprehen-*, *understand-* and its derivatives, *wit* and its derivatives, *idio-* (without *idiom*), *ideot*, *stupid-* ("smart" or "fool" are too polysemous to be taken into account, as their concordance shows). *Emma* also comes first for this lexical field in the Reference corpus, where the number of occurrences making up the list is similar to that in WOMEN (respectively 4014 and 4026).

⁶¹ With 2 uses of "reader" *Rachel* is a positive user but does not stand out in WOMEN. The plural appears only once, in the preface. *Betsy Thoughtless* towers above all other texts for its use of "reader" both in the Women and the Reference corpus: 64 of the 180 occurrences of "reader" in WOMEN rather than the 8 expected. As the presence of "reader" in Haywood's fiction mainly occurs in *Betsy Thoughtless* and *Jemmy & Jenny Jessamy*, rather than in her earlier texts, this can be seen as a feature of the "new species of writing" introduced by Fielding and dwindling after Sterne.

⁶² In WOMEN, 14 occurrences of "be remembered" which implies the reader, along with 2 of "the reader may remember," with a concentration of 5 "may be remembered" in *Stella* and 4 "it will be remembered" in *Tynemouth*: authors who use this formula to refresh their reader's memory tend to do so repetitively. Eliza Haywood is an earlier example of this.

⁶³ Rachel calls him "a prosing disagreeable man" (42) before she learns to appreciate him. Tomkins is in this respect not unlike More's "unlikable prig" Coelebs – although it must be noted that the latter judgment is not that of the author but of an editor (Demers 27).

⁶⁴ "'I wish we were not going this evening,' said Elizabeth; 'they say that Mrs. Fellows is so clever, and so satirical, that I shall be afraid of speaking a word.'" To which "Emily" replies "Dear now, I am glad we are going" (*Display* 1). The two protagonists are then referred to as "these young people" and contrasted, "belong[ing] to opposite classes of character" (2), not unlike Rachel and Sophia. Thanks are due to Darren Bevin for checking from the Chawton collection that *Prudence and Principle* opens differently.

⁶⁵ "—You have been in France? said my gentleman" (Sterne, *A Sentimental Journey* 3).

⁶⁶ Taylor's readers may have been more familiar with *Beauties of Sterne* than with *Tristram Shandy*. She refers openly to Sterne in *The Authoress* when the aspiring writer gives a chapter full of dashes and incomplete sentences and snaps at her stern reader, "Did you ever hear of Sterne [...] in a tone which blended contempt for his ignorance, and mortification at his not having made the discovery without her assistance" (8-9). This triggers "It is a difficult and dangerous thing to affect the style of any author, particularly one so peculiar as that of Sterne. I have seen other attempts lately, but they all fail" (16). "Alas poor Rachel" quoted earlier also recalls Sterne and widens intertextuality to Shakespeare.

⁶⁷ The Open University's fancy but unreliable search engine mentioned earlier spots only 9 occurrences as two of them are spread over two lines.

⁶⁸ One of these texts is from 1769 (the earlier mentioned *The Reward of Virtue; or the History of Miss Polly Graham*), the other from 1812 (Elizabeth Helme's [Magdalen, or the Penitent of Godstow](#)) and 1823 ([Isabella](#)), so that chronology does not seem an issue for the complete absence of the word "sensibility". Jacson, the author of *Isabella*, is not averse to the term as it appears twice in her 1812 [Things by their Right Name](#).

⁶⁹ Of course the choice of texts plays a role here, but the diminishing use of the term appears clearly in both corpora.

⁷⁰ *Sense and Sensibility* only uses the second of these two words 10 times in slightly more than four times as much text: Austen's novel has about 144 000 words, *Rachel* 34,000.

⁷¹ "Contemporary Reviews," *British Fiction, 1800-1829*.

⁷² This appears in the comparison of *Rachel* with texts of comparable size in part-of-speech tagged versions (Tree-Tagger and TXM): *The Man of Feeling*, *Polly*, Wollstonecraft's *Mary and Maria*, Opie's *A Woman's Love* (admittedly a small sample of texts – but this is for the time being an experimental corpus). *Hyperbase*'s "positive specificity" list confirms the proportionally large amount of names. *Rachel*'s centrality for the use of tokens in the Women corpus results from the paucity of "specific" terms other than names.

PRIMARY SOURCES

Ashton Priory. 3 vols. London: Minerva, 1792. [Chawton Novels Online](#). Web. 3. Oct. 2015.

Austen, Jane. *Emma*. 1816. *Project Gutenberg*. Web. 3 Oct. 2015.
<<https://www.gutenberg.org/ebooks/158>>.

---. *Jane Austen's Letters*. Ed. Deirdre Le Faye. 4th ed. Oxford: Oxford UP, 2011. Print.

---. *Mansfield Park*. 1814. *Project Gutenberg*. Web. 3 Oct. 2015.
<<https://www.gutenberg.org/ebooks/141>>.

---. *Pride and Prejudice*. 1813. *The Republic of Pemberley*. Web. 3 Oct. 2015.

<<http://pemberley.com/janeinfo/janewrit.html#janovls>>.

Bennett, Anna Maria. *Agnes De-Courci: A Domestic Tale*. 4 vols. Bath: Hazard, 1789. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Brunton, Mary. *Self-Control*. 1810. *Project Gutenberg*. Web. 3 Oct. 2015.
<<https://www.gutenberg.org/files/41196/41196-h/41196-h.htm>>.

Burney, Frances. *Cecilia; Or Memoirs of a Heiress*. 1782. 3 vols. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/6346>,
<https://www.gutenberg.org/ebooks/7146>, <https://www.gutenberg.org/ebooks/7152>>.

---. *Evelina; Or, the History of a Young Lady's Entrance into the World*. 1778. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/6053>>.

---. *The Wanderer; or, Female Difficulties*. 5 vols. 1814. *Project Gutenberg*. Web. 3 Oct. 2015.
<<https://www.gutenberg.org/ebooks/37437>, <https://www.gutenberg.org/ebooks/37438>,
<https://www.gutenberg.org/ebooks/37439>, <https://www.gutenberg.org/ebooks/37440>,
<https://www.gutenberg.org/ebooks/37441>>.

Canning, Mary Ann. *The Offspring of Fancy*. 2 vols. London: Bew, 1778. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Carver, Mrs. *The Old Woman*. London: Minerva, 1800. [Chawton Novels Online](#). Web.

The Cautious Lover; or, The History of Lord Woburn. 2 vols. London: Davies, 1772. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Cooper, Maria Susanna. *The Wife; or, Caroline Herbert*. 2 vols. London: Becket & Porter, 1813. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Craik, Helen. *Stella of the North, or the Foundling of the Ship*. London: Minerva, 1802. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Disraeli, Benjamin. *Vivian Grey*. 1826. New York: Century, 1906. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.Project Gutenberg. Web. org/files/9840/9840-h/9840-h.htm>>.

Edgeworth, Maria. *Belinda*. 1802. *Tales and Novels*, Vol. 3. London: Simpkin & Marshall, 1857. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/9455>>.

---. *Helen*. 1834. *Tales and Novels*, Vol. 10. London: Simpkin & Marshall, 1857. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/8531>>.

Ferrier, Susan. *Marriage*. 1818. London: Bentley, 1881. *Project Gutenberg*. Web. 3 Oct. 2015.
<<https://www.gutenberg.org/ebooks/12669>>.

- Foster, Mrs. E.M. *The Corinna of England, and a Heroine in the Shade: A Modern Romance*. 2 vols. London: Crosby, 1809. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- . *Substance and Shadow; or, the Fisherman's Daughters of Brighton*. London: Minerva, 1812. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Galt, John. *The Ayrshire Legatees; Or, The Pringle Family*. 1821. London: Macmillan, 1895. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/1384>>.
- . *The Provost*. 1822. London: Foulis, 1913. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/1296>>.
- Godwin, William. *Caleb Williams; Or, Things as They Are*. 1794. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/11323>>.
- . *Damon & Delia: A Tale*. London: Hookham, 1784. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/10318>>.
- Goldsmith, Oliver. *The Vicar of Wakefield*. 1766. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/2667>>.
- Green, Sarah. *Romance Readers and Romance Writers: A Satirical Novel*. 3 vols. London: Hookham & Hookham, 1810. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Harvey, Jane. *Any Thing But What You Expect*. 3 vols. Derby: Mozley, 1819. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- . *The Castle of Tynemouth. A Tale*. 2 vols. Newcastle upon Tyne: MacKenzie, 1830. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Hatton, Anne Julia Kemble. *Lovers and Friends; or, Modern Attachments*. 3 vols. London: Minerva, 1821. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Hays, Mary. *The Memoirs of Emma Courtney*. 1796. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/41256>>.
- Haywood, Eliza. *The History of Miss Betsy Thoughtless*. 1752. *Eighteenth-century Fiction Full Text Database*. Alexandria, VA: Chadwick-Healey, 1996. CD-ROM.
- Helme, Elizabeth. *Magdalen; or, the Penitent of Godstow*. 3 vols. Brentford: Norbury, 1812. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- The History of Lord Clayton and Miss Meredith*. 2 vols. London: Robinson & Roberts, 1771. [Chawton Novels Online](#). Web. 3 Oct. 2015.

- Hogg, James. *The Private Memoirs and Confessions of a Justified Sinner*. 1824. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/2276>>.
- . *The Three Perils of Man*. London: Longman, 1822. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/39776>>.
- Holcroft, Thomas. *The Adventures of Hugh Trevor*. 1794. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/9811>>.
- . *Anna St Ives*. 1792. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/9468>>.
- Hughes, Anne. *Caroline; or, the Diversities of Fortune*. 3 vols. London: Minerva, 1787. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Hunter, Rachel. *The Unexpected Legacy*. 2 vols. London: Longman & Robberds, 1804. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Inchbald, Elizabeth. *Nature and Art*. 1796. London: Cassell, 1886. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/3787>>.
- . *A Simple Story*. 1791. 4th ed. London: Robinson, 1799. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/22002>>.
- Jacson, Frances. *Isabella, a Novel*. 3 vols. London: Colburn, 1823. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- . *Things By Their Right Names*. 2 vols. London: Robinson, 1812. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Johnson, Mrs. Francis. *the Philanthropist: An Unfashionable Tale*. Dublin: White, 1786. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Lewis, Alethea. *Vicissitudes in Genteel Life*. 4 vols. Stafford: Longman, 1794. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Mackenzie, Anna Maria. *The Irish Guardian, or, Errors of Eccentricity*. 3 vols. London: Longman, Hurst, Rees & Orme, 1809. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Mackenzie, Henry. *The Man of Feeling*. 1771. London: Cassell, 1886. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/5083>>.
- Martin, Mrs. *The Enchantress; or, Where Shall I Find Her? A Tale*. London: Minerva, 1801. [Chawton Novels Online](#). Web. 3 Oct. 2015.

- Matthews, Mrs. *Simple Facts or the History of an Orphan*. 3 vols. London: printed by Low for the author, 1793. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- More, Hannah. *Coelebs in Search of a Wife*. 1809. New York: Derby & Jackson, 1858. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/31879>>.
- Opie, Amelia. *Adeline Mowbray; or, The Mother and Daughter*. 1804. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/37908>>.
- Peacock, Thomas Love. *Headlong Hall*. 1815. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/12803>>.
- Prepossession; or, Memoirs of Count Touloussin*. 2 vols. London: Forbes, Hookham & Carpenter, 1792. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Purbeck, Elizabeth and Jane Purbeck. *Honoria Sommerville*. 4 vols. London: Robinson, 1789. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- The Reward of Virtue; or, the History of Miss Polly Graham*. London: Roson & Cooke, 1769. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Scott, Walter. *The Antiquary*. 1816. Boston: Estes and Lauriat, 1893. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/7005>>.
- . *St Ronan's Well*. 1823. Boston: Estes and Lauriat, 1894. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/20749>>.
- Shelley, Mary. *Frankenstein, or the Modern Prometheus*. 1831. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/84>>.
- Smith, Charlotte Turner. *Emmeline, the Orphan of the Castle*. 1788. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/41646>>.
- Smollett, Tobias. *The Adventures of Roderick Random*. 1748. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/4085>>.
- . *The Expedition of Humphrey Clinker*. 1771. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/2160>>.
- Spence, Elizabeth Isabella. *The Curate and His Daughter, a Cornish Tale*. 3 vols. London: Longman, Hurst, Rees, Orme & Brown, 1813. [Chawton Novels Online](#). Web. 3 Oct. 2015.
- Sterne, Laurence. *A Sentimental Journey*. 1768. *Project Gutenberg*. Web. 3 Oct. 2015.

<<https://www.gutenberg.org/ebooks/804>>.

The Reward of Virtue; or the History of Miss Polly Graham. London: Roson, Pyne and Cooke, 1769. [Chawton Novels Online](#). Web. 3 Oct. 2015.

The Suspicious Lovers. 3 vols. London: Wilkie & Dilly, 1777. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Taylor, Jane. *Rachel: A Tale*. London: Taylor & Hessey, 1817. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Tomlins, Elizabeth Sophia. *Tomlins, the Victim of Fancy*. 2 vols. London: Baldwin & Wilkie, 1787. [Chawton Novels Online](#). Web. 3 Oct. 2015.

The Village Coquette. 3 vols. London: Whittaker, 1822. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Three Weeks in the Downs, or Conjugal Fidelity Rewarded: Exemplified in the Narrative of Helen and Edmund. London: Bennet, 1829. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Wilkinson, Sarah Scudgell. *The Child of Mystery*. 3 vols. London: Hughes, 1808. [Chawton Novels Online](#). Web. 3 Oct. 2015.

Wollstonecraft, Mary. *Mary: A Fiction*. London: Johnson, 1788. *Project Gutenberg*. Web. 3 Oct. 2015. <<https://www.gutenberg.org/ebooks/16357>>.

Works Cited

Allison, Sarah, et al. "Quantitative Formalism: An Experiment." *Literary Lab Pamphlet 1*, Stanford, 15 Jan. 2011. Web. 3 Oct. 2015. <<https://litlab.stanford.edu/pamphlets/>>.

Bandry-Scubbi, Anne. "Chawton House Library Novels Online." *The Female Spectator* 18.3 (2014): 12-13. Print.

---. "Evelina, Lydia, Isabella: Quoi de neuf dans le (nouveau) Roman Anglais féminin à la lumière d'une analyse de corpus." *RSEAA XVII-XVIII HS3* (2013): 113-31. Print.

---. "Gulliver et la machine à compter: Une étude de spécificités." *BSEAA XVII-XVIII* 53.1 (2001): 145-57. Web. 3 Oct. 2015. <http://www.persee.fr/web/revues/home/prescript/article/xvii_0291-3798_2001_num_53_1_1601>.

---. "Les livres de Sterne: Suites et fins." *BSEAA XVII-XVIII* 50.1 (2000): 115-36. Web. 3 Oct. 2015. <http://www.persee.fr/web/revues/home/prescript/article/xvii_0291-3798_2000_num_50_1_1481>.

-
- . "Renaissance de/chez Eliza Haywood." *Littérature Anglo-saxonne au féminin: (Re)naissance(s) et horizons, XVIIIe-XXe siècles*. Ed. Claire Bazin and Guyonne Leduc. Paris: L'Harmattan (Des idées et des femmes), 2012. 17-39. Print.
- . "Roderick Random amidst Eighteenth-Century Fiction: A Computer-aided Textual Analysis." *RSEAA XVII-XVIII* 66.1 (2009): 205-25. Web. 3 Oct. 2015. <http://www.persee.fr/web/revues/home/prescript/article/xvii_0291-3798_2009_num_66_1_2399>.
- . "Space and Emotions: The Use of Toponyms in *Pride and Prejudice*." *Georgian Cities*. 2014. Web. 3 Oct. 2015. <<http://www.18thc-cities.paris-sorbonne.fr/Space-and-Emotions?lang=en>>.
- . "Yes, Novels: *Evelina*, *Cecilia* and *Belinda*, or the Beginning of the Novel of Manners Revisited." *Études Anglaises* 2.66 (2013): 234-49. Web. 3 Oct. 2015. <<http://www.cairn.info/publications-de-Bandry-Scubbi-Anne--108693.htm>>.
- Barker-Benfield, G.J. *The Culture of Sensibility: Sex and Society in Eighteenth-Century Britain*. Chicago: U of Chicago P, 1992. Print.
- Biber, Douglas. "Corpus Linguistics and the Study of Literature: Back to the Future?" *Scientific Study of Literature* 1.1 (2011): 15-23. Web. 14 July 2014. <<http://www.jbe-platform.com/content/journals/10.1075/ssol.1.1.02bib>>.
- Blakey, Dorothy. *The Minerva Press: 1790-1820*. Oxford: Oxford UP, 1939. Print.
- Brewer, David A. "Counting, Resonance, and Form, A Speculative Manifesto (with Notes)." *Eighteenth-Century Fiction* 24.2 (Winter 2011-12): 161-70. Print.
- British Fiction, 1800-1829: A Database of Production, Circulation & Reception*. Cardiff: Centre for Editorial and Intertextual Research. 2004. Web. 3 Oct. 2015. <<http://www.british-fiction.cf.ac.uk/>>.
- Brown, Susan, Patricia Clements, and Isobel Grundy, ed. *Orlando: Women's Writing in the British Isles from the Beginnings to the Present*. Cambridge: Cambridge U P Online, 2006. Web. 3 Oct. 2015. <<http://orlando.cambridge.org/>>.
- Brunet, Etienne. *Manuel de référence pour Hyperbase 9.0*. Nice: Université de Nice Sophia-Antipolis, 2011. Print.
- Burrows, John. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon, 1987. Print.
- . "Never Say Always Again: Reflections on the Number Game." *Text and Genre in Reconstruction*. Ed. Willard McCarty. Cambridge: Open Book, 2010. 13-36. Print.

- . "Style." *The Cambridge Companion to Jane Austen*. Ed. Edward Copeland and Juliet McMaster. Cambridge: Cambridge UP, 1997. 170-88. Print.
- Butler, Marilyn. *Jane Austen and the War of Ideas*. Oxford: Clarendon, 1975. Print.
- Chawton House Library*. Web. 2 Oct. 2015.
- Corman, Brian. *Women Novelists before Jane Austen: The Critics and Their Canons*. Toronto: U of Toronto P, 2008. Print.
- Cossette, André. *La Richesse lexicale et sa mesure*. Paris: Champion, 1994. Print.
- Dallas, George. "Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension Reduction." 30 Oct. 2013. Web. 3 Oct. 2015.
<<https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/>>.
- Demers, Patricia. Introduction. *Cælebs in Search of a Wife*. By Hannah More. 1809. Peterborough, ON: Broadview, 2007. Print.
- Dixon, Thomas. "Enthusiasm Delineated: Weeping as a Religious Activity in Eighteenth-Century Britain." *Litteraria Pragensia: Studies in Literature and Culture* 22.43 (2013): 59-81. Web. 3 Oct. 2015..
- . *Weeping Britannia: Portrait of a Nation in Tears*. Oxford: Oxford UP, forthcoming. Print.
- Fischer-Starcke, Bettina. *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*. London: Continuum, 2010. Print.
- Garside, Peter. "The English Novel in the Romantic Era: Consolidation and Dispersal." *The English Novel 1770-1829*. Ed. Peter Garside, James Raven, and Rainer Schöwerling. Vol. 2. Oxford: Oxford UP, 2000. 15-103. Print.
- Gemmett, Robert J. Introduction. *Modern Novel Writing, or the Elegant Enthusiast*. By William Beckford. 1796. Stroud: Nonsuch, 2008. Print.
- Graves, David Andrew. "Computer Analysis of Word Usage in *Emma*." *Persuasions* 21 (1999): n. pag. Web. 3 Oct. 2015..
<<http://www.jasna.org/persuasions/printed/number21/index.html>>.
- . "Vocabulary Profiles of Letters and Novels of Jane Austen and her Contemporaries." *Persuasions* 26.1 (2005): n. pag. Web. 3 Oct. 2015. .
<<http://www.jasna.org/persuasions/on-line/vol26no1/graves.htm>>.

- Grundy, Isobel. "Chawton House: Gathering Old Books for a New Library." *British Women's Writing in the Long Eighteenth Century: Authorship, Politics and History*. Ed. Jenny Batchelor and Cora Kaplan. Basingstoke: Palgrave Macmillan, 2005. 171-85. Print.
- . "(Re)discovering Women's Texts." *Women and Literature in Britain 1700-1800*. Ed. Vivien Jones. Cambridge: Cambridge UP, 2000. 179-96. Print.
- Heuser, Ryan, and Long Le-Khac. "A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method." *Literary Lab Pamphlet 4* Stanford, May 2012. Web. 3 Oct. 2015 <<https://litlab.stanford.edu/pamphlets>>.
- Hinton, Perry R. *Statistics Explained*. London: Routledge, 2004. Print.
- Ho, Yufang. *Corpus Stylistics in Principles and Practice: A Stylistic Exploration of John Fowles' The Magus*. London: Continuum, 2011. Print.
- HYPERBASE: Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels*. Version 9. 2011. Université de Nice. Web. 3 Oct. 2015.
- Jockers, Matthew L. *Macroanalysis: Digital Methods & Literary History*. Urbana, Chicago: U of Chicago P, 2013. Print.
- Johnson, Samuel. *The Rambler 4. The Oxford Authors: Samuel Johnson*. Ed. Donald Greene. Oxford: Oxford UP, 1984. 175-79. Print.
- Kelly, Gary. *English Fiction of the Romantic Period 1789-1830*. London: Longman, 1990. Print.
- Koppel, Moshe, Shlomo Argamon, and Anat Rachel Shimoni. "Automatically Categorizing Written Texts by Author Gender." *Literary and Linguistic Computing* 17.4 (2002):401-12. Print.
- Kulik, Maggie. "What the Bookseller Did: A Case of Eighteenth-Century Plagiarism." *The Female Spectator* 2000. Print.
- Labbé, Dominique. *Corneille dans l'ombre de Molière*. Paris: Impressions Nouvelles, 2003. Print.
- Labbe, Jacqueline M., ed. *The History of British Women's Writing, 1750-1830*. Houndsmill Basingstoke: Palgrave Macmillan, 2010. Print.
- Lebart, L. and A Salem. *Statistique textuelle*. Paris: Dunod, 1994. Print.
- Mahlberg, Michaela. "Corpus Stylistics: Bridging the Gap between Linguistic and Literary Studies." *Text, Discourse and Corpora: Theory and Analysis*. Ed. Michael Hoey, et al. London: Continuum, 2007. 219-44. Print.

-
- . *Corpus Stylistics and Dickens's Fiction*. London: Routledge, 2013. Print.
- McLeod, Deborah Anne. *The Minerva Press*. Diss. U of Alberta, 1997. Web. 3 Oct. 2015. <<https://era.library.ualberta.ca/public/view/item/uuid:44c957fb-bf42-4aa8-a297-ef40f30719eb/>>.
- Milic, Louis T. *A Quantitative Approach to the Style of Jonathan Swift*. The Hague: Mouton, 1967. Print.
- Moretti, Franco. *Distant Reading*. London: Verso, 2013. Print.
- Moretti, Franco, and Mark Algee-Hewitt. *Stanford Literary Lab*. Stanford U, 2015. Web. 5 Sept. 2015.
- Mosteller, Frederick, and David Lee Wallace. *Inference and Disputed Authorship: The Federalist*. Reading, Ma.: Addison-Wesley, 1964. Print.
- Muller, Charles. *Essai de statistique lexicale: L'illusion comique de Pierre Corneille*. Paris: Klincksieck, 1964. Print.
- Murphy, Olivia. "Jane Austen's Critical Response to Women's Writing: 'A Good Spot for Fault-Finding'." Labbe. 288-300.
- Oates, J.C.T. *Shandyism & Sentiment 1760-1800*. Cambridge: Cambridge Bibliographical Society, 1968. Print.
- Oxford English Dictionary*. 2015. *Oxford English Dictionary*. Web. 3 Oct. 2015.
- Probyn, Clive. *English Fiction of the 18th Century, 1700-1789*. London: Longman, 1987. Print.
- Rastier, François. *La Mesure et le grain: Sémantique de corpus*. Paris: Champion, 2011. Print.
- Raven, James. "Historical Introduction: The Novel Comes of Age." *The English Novel 1770-1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*. Ed. Peter Garside, James Raven, and Rainer Schöwerling. Vol. 1. Oxford: Oxford UP, 2000. 15-121. Print.
- Spacks, Patricia Meyer. *Novel Beginnings: Experiments in Eighteenth-Century English Fiction*. New Haven: Yale UP, 2006. Print.
- Sterne, Laurence. *The Life and Opinions of Tristram Shandy, Gentleman*. Vol. 1. 1759. London: Dodsley, 1760. Web. 3 Oct. 2015. <<http://www.tristramshandyweb.it/>>.
- . *A Sentimental Journey*. 1768. Indianapolis: Hackett, 2006. Print.

Taylor, Jane. *The Authoress. A Tale*. London: Taylor and Hessey, 1819. Print.

---. *Display: A Tale*. London: Taylor and Hessey, 1815. Print.

---. *Prudence and Principle*. London: Taylor and Hessey, 1821. Print.

Tirvengadam, Vina. "Linguistic Fingerprints and Literary Fraud." *CHWP* U of Manitoba, 1998. Web. 3 Oct. 2015. <<http://projects.chass.utoronto.ca/chwp/tirven/>>.

Todd, Janet. *Sensibility: An Introduction*. London: Methuen, 1986. Print.

Tompkins, J.M.S. *The Popular Novel in England 1770-1800*. London: Methuen, 1932. Print.

Waldron, Mary. "The Frailties of Fanny: *Mansfield Park* and the Evangelical Movement." *Eighteenth-Century Fiction* 6.3.5 (1994) n. pag. Web. 3 Oct. 2015.

Yeazell, Ruth Bernard. *Fictions of Modesty: Women and Courtship in the English Novel*. Chicago: U of Chicago P, 1991. Print.